ERGODIC AVERAGES OF COMMUTING TRANSFORMATIONS WITH DISTINCT DEGREE POLYNOMIAL ITERATES

QING CHU, NIKOS FRANTZIKINAKIS, AND BERNARD HOST

ABSTRACT. We prove mean convergence, as $N \to \infty$, for the multiple ergodic averages $\frac{1}{N} \sum_{n=1}^{N} f_1(T_1^{p_1(n)}x) \cdot \ldots \cdot f_\ell(T_\ell^{p_\ell(n)}x)$, where $p_1, \ldots, p_\ell$ are integer polynomials with distinct degrees, and $T_1, \ldots, T_\ell$ are commuting, invertible measure preserving transformations, acting on the same probability space. This establishes several cases of a conjecture of Bergelson and Leibman, that complement the case of linear polynomials, recently established by Tao. Furthermore, we show that, unlike the case of linear polynomials, for polynomials of distinct degrees, the corresponding characteristic factors are mixtures of inverse limits of nilsystems. We use this particular structure, together with some equidistribution results on nilmanifolds, to give an application to multiple recurrence and a corresponding one to combinatorics.

Contents

1. Main results, ideas in the proofs, and further directions

1.1. **Introduction and main results.** A well studied and difficult problem in ergodic theory is the analysis of the limiting behavior of multiple ergodic averages of commuting transformations taken along polynomial iterates. A related conjecture of Bergelson and Leibman (given explicitly in [6]) states the following:

Conjecture. *Let $(X, \mathcal{X}, \mu)$ be a probability space, $T_1, \ldots, T_\ell \colon X \to X$ be commuting, invertible measure preserving transformations, $f_1, \ldots, f_\ell \in L^\infty(\mu)$, and $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$.*

Then the limit

$$(1) \qquad \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f_1(T_1^{p_1(n)}x) \cdot \ldots \cdot f_\ell(T_\ell^{p_\ell(n)}x)$$

exists in $L^2(\mu)$.

Special forms of the averages in (1) were introduced and studied by Furstenberg [19], Furstenberg and Katznelson [20], and Bergelson and Leibman [8], in a depth that was sufficient for them to establish the theorem of Szemerédi on arithmetic progressions and its multidimensional and polynomial extensions respectively.

Proving convergence of these averages turned out to be a harder problem. When all the transformations $T_1, \ldots, T_\ell$ are equal, convergence was established after a long series of intermediate results; the papers [19, 11, 12, 13, 21, 30, 23, 34] dealt with the important case of linear polynomials, and using the machinery introduced in [23], convergence for arbitrary polynomials was finally obtained in [24] except for a few cases that were treated in [28]. For general commuting transformations, progress has been scarcer. When all the polynomials in (1) are linear, after a series of partial results [20, 11, 29, 33, 16] that were obtained using ergodic theory, convergence was established in [31] using a finitary argument. Subsequently, motivated by ideas from [31], several other proofs of this "linear" result were found using non-standard analysis [32], and then ergodic theory [2, 22]. Proofs of convergence for general polynomial iterates have been given only under very strong ergodicity assumptions [5, 27]. On the other hand, very recently, in [3, 4] techniques from [2] have been refined and extended, aiming to eventually handle the case of general polynomial iterates. Despite such intense efforts, for general commuting transformations, apart from the case where all the polynomials are linear, no other instance of the conjecture of Bergelson and Leibman has been resolved. In this article, we are going to establish this conjecture when the polynomial iterates have distinct degrees:

Theorem 1.1. *Let $(X, \mathcal{X}, \mu)$ be a probability space, $T_1, \ldots, T_\ell \colon X \to X$ be commuting, invertible measure preserving transformations, and $f_1, \ldots, f_\ell \in L^\infty(\mu)$. Suppose that the polynomials $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$ have distinct degrees.*

Then the limit

$$(2) \qquad \lim_{N-M \to \infty} \frac{1}{N-M} \sum_{n=M}^{N-1} f_1(T_1^{p_1(n)}x) \cdot \ldots \cdot f_\ell(T_\ell^{p_\ell(n)}x)$$

exists in $L^2(\mu)$.

Unlike previous arguments in [11, 13, 31, 32, 2, 22], where one finds ways to sidestep the problem of giving precise algebraic descriptions of the factor systems that control the limiting behavior of special cases of the averages (2), a distinctive feature of the proof of Theorem 1.1 is that we give such descriptions.[1] Furthermore, we did not find it advantageous to work within a suitable extension of our system in order to simplify our study (like the "pleasant" or "magic" extensions that were introduced in [2] and in [22] respectively). In this respect, our analysis is

---

[1]A key difference between the averages of $f_1(T_1^n x) \cdot f_2(T_2^n x)$ and the averages of $f_1(T_1^n x) \cdot f_2(T_2^{n^2} x)$ is that when $T_1 = T_2$ the first one becomes "degenerate" (= averages of $(f_1 \cdot f_2)(T_1^n x)$), and this complicates the structure of the possible factors that control their limiting behavior. However, no such choice of T_1, T_2 makes the second average "degenerate".

more closely related to the one used to study convergence results when all the transformations $T_1, \ldots, T_\ell$ are equal, and in fact uses this well developed single transformation theory in a crucial way (in some special cases our approach leads to very concise proofs, see Appendix A). The next result gives the description of the aforementioned factors (the factors $\mathcal{Z}_{k,T_i}$ are defined in Section 2.2):

Theorem 1.2. *Let $(X, \mathcal{X}, \mu)$ be a probability space, $T_1, \ldots, T_\ell \colon X \to X$ be commuting, invertible measure preserving transformations, and $f_1, \ldots, f_\ell \in L^\infty(\mu)$. Let $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$ be polynomials with distinct degrees and maximum degree d.*
   Then there exists $k = k(d, \ell) \in \mathbb{N}$ such that: If $f_i \perp \mathcal{Z}_{k,T_i}$ for some $i \in \{1, \ldots, \ell\}$, then

$$\lim_{N-M \to \infty} \frac{1}{N-M} \sum_{n=M}^{N-1} f_1(T_1^{p_1(n)} x) \cdot \ldots \cdot f_\ell(T_\ell^{p_\ell(n)} x) = 0$$

in $L^2(\mu)$.

Factors that satisfy the aforementioned convergence property are often called *characteristic factors*. The utility of the characteristic factors obtained in Theorem 1.2 stems from the fact that each individual factor is a mixture of systems of algebraic origin, in particular, it is a mixture of inverse limits of nilsystems [23] (see also Theorem 2.1). Using this algebraic description of the characteristic factors (in fact its consequence Proposition 3.1 is more suitable for our needs), and some equidistribution results on nilmanifolds, we give the following application to multiple recurrence:

Theorem 1.3. *Let $(X, \mathcal{X}, \mu)$ be a probability space and $T_1, \ldots, T_\ell \colon X \to X$ be commuting, invertible measure preserving transformations.*
   Then for every choice of distinct positive integers $d_1, \ldots, d_\ell$, and every $\varepsilon > 0$, the set

$$(3) \qquad \{n \in \mathbb{N} \colon \mu(A \cap T_1^{-n^{d_1}} A \cap \cdots \cap T_\ell^{-n^{d_\ell}} A) \geq \mu(A)^{\ell+1} - \varepsilon\}$$

has bounded gaps.

If the integers are not distinct, say $d_1 = d_2$, then the result fails. For example, one can take $T_2 = T_1^2$, and choose the (non-ergodic) transformation T_1, and the set A, so that (3) fails with any power of $\mu(A)$ on the right hand side for every $n \in \mathbb{N}$ (Theorem 2.1 in [7]). If $\ell = 2$, $d_1 = d_2 = 1$, and the joint action of the transformations T_1, T_2 is ergodic, then the result remains true up to a change of the exponent on the right hand side [10]. But even under similar ergodicity assumptions, the result probably fails when 3 exponents agree no matter what exponent one uses on the right hand side (a conditional counterexample appears in Proposition 5.2 of [14]).

It will be clear from our argument that in the statement of Theorem 1.3 we can replace the polynomials $n^{d_1}, \ldots, n^{d_\ell}$ by any collection of polynomials $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$ with zero constant terms that satisfy $t^{\deg(p_i)+1} | p_{i+1}$ for $i = 1, \ldots, \ell - 1$. For example $\{n, n^3 + n^2, n^5 + n^4\}$ is such a family. On the other hand, our argument does not work for all polynomials with distinct degrees (the problem is to find a replacement for Lemma 7.6), but the same lower bounds are expected to hold for any collection of rational independent integer polynomials with zero constant terms.

Using a multidimensional version of Furstenberg's correspondence principle (see [20] or [8]) it is straightforward to give a combinatorial consequence of this result. We leave the routine details of the proof to the interested reader.

Theorem 1.4. *Let $k, \ell \in \mathbb{N}$, $\Lambda \subset \mathbb{Z}^k$ with $\bar{d}(\Lambda) > 0$,[2] and $v_1, \ldots, v_\ell$ be vectors in $\mathbb{Z}^k$.*
Then for every choice of distinct positive integers $d_1, \ldots, d_\ell$, and every $\varepsilon > 0$, the set

$$(4) \qquad \{n \in \mathbb{N} \colon \quad \bar{d}\big(\Lambda \cap (\Lambda + n^{d_1} v_1) \cap \cdots \cap (\Lambda + n^{d_\ell} v_\ell)\big) \geq \bar{d}(\Lambda)^{\ell+1} - \varepsilon\}$$

has bounded gaps.

1.2. Ideas in the proofs of the main results.

1.2.1. *Key ingredients.* The proofs of Theorems 1.1, 1.2, and 1.3, use several ingredients.

Van der Corput's Lemma. We are going to use repeatedly the following variation of the classical elementary lemma of van der Corput. Its proof is a straightforward modification of the one given in [5].

Van der Corput's Lemma. *Let $\{v_n\}_{n \in \mathbb{N}}$ be a bounded sequence of vectors in a Hilbert space. Let*

$$b_h = \overline{\lim}_{N-M \to \infty} \Big| \frac{1}{N-M} \sum_{n=M}^{N-1} <v_{n+h}, v_n> \Big|.$$

Suppose that

$$\lim_{H \to \infty} \frac{1}{H} \sum_{h=1}^{H} b_h = 0.$$

Then

$$\lim_{N-M \to \infty} \Big\| \frac{1}{N-M} \sum_{n=M}^{N-1} v_n \Big\| = 0.$$

In most applications we have $b_h = 0$ for every sufficiently large h, or for "almost every h", meaning that the exceptional set has zero upper density.

An approximation result. This enables us in several instances to replace sequences of the form $(f(T^n x))_{n \in \mathbb{N}}$, where f is a $\mathcal{Z}_{k,T}$-measurable function, with k-step nilsequences. For ergodic systems, this result is an easy consequence of the structure theorem for the factors $\mathcal{Z}_{k,T}$ (Theorem 2.1). But we need a harder to establish non-ergodic version (Proposition 3.1); in our context we cannot assume that each individual transformation is ergodic.

Nilsequence correlation estimates. These, roughly speaking, assert that "uniform" sequences do not correlate with nilsequences (see for example Theorem 6.2).

Equidistribution results on nilmanifolds. These will only be used in the proof of Theorem 1.3 (see Section 7.1).

---

[2]For a set $\Lambda \subset \mathbb{Z}^k$, we define its upper density by $\bar{d}(\Lambda) = \limsup_{N \to \infty} |\Lambda \cap [-N, N]^k| / (2N)^k$.

1.2.2. *Combining the key ingredients.* We first prove Theorem 1.2 that provides convenient characteristic factors for the multiple ergodic averages in (2). Its proof proceeds in two steps: (i) In Sections 4 and 5 we use a PET-induction argument based on successive uses of van der Corput's Lemma to find a characteristic factor for the transformation that corresponds to the highest degree polynomial iterate, and (ii) In Section 6 we combine step (i), with the aforementioned approximation result and nilsequence correlation estimates, to find characteristic factors for the other transformations as well.

The strategy for proving Theorems 1.1 and 1.3 can be summarized as follows: In order to study the limit (2), we first use Theorem 1.2 to reduce matters to the case where all the functions f_i are $\mathcal{Z}_{k,T_i}$-measurable for some $k \in \mathbb{N}$, and then the aforementioned approximation result to reduce matters to establishing certain convergence or equidistribution properties on nilmanifolds. This last step is easy to carry out when proving Theorem 1.1 (see Section 6.4), but becomes much more cumbersome when proving Theorem 1.3. We prove the equidistribution properties needed for Theorem 1.3 in Section 7.

1.3. **Further directions.** When $\ell = 2$ and $p_1(n) = p_2(n) = n$, it is known that some sort of commutativity assumption on the transformations T_1, T_2 has to be made in order for the limit (2) to exist in $L^2(\mu)$ (see [9] for examples where convergence fails when T_1 and T_2 generate solvable groups of exponential growth). On the other hand, it is not clear whether a similar assumption is necessary when say $\ell = 2$ and $p_1(n) = n$, $p_2(n) = n^2$. In fact, it could be the case that for Theorems 1.1, 1.2, and 1.3, no commutativity assumption at all is needed.

Since convergence of the averages in (2) for $\mathcal{Z}_k$-measurable functions can be shown for general families of integer polynomials (see the argument in Section 6.5), it follows that the averages in (2) converge in $L^2(\mu)$ for any collection of polynomials for which the conclusion of Theorem 1.2 holds. We conjecture that the conclusion of Theorem 1.2 holds if and only if the family of polynomials $p_1, \ldots, p_\ell$ is pairwise independent, meaning, the set $\{1, p_i, p_j\}$ is linearly independent for every $i, j \in \{1, \ldots, \ell\}$ with $i \neq j$ (simple examples show that the condition is necessary). Furthermore, we conjecture that if the polynomials $1, p_1, \ldots, p_\ell$ are linearly independent, then the factors $\mathcal{K}_{\mathrm{rat}}(T_i)$ can take the place of the factors $\mathcal{Z}_{k,T_i}$ in the hypothesis of Theorem 1.2.

In the case where all the transformations $T_1, \ldots, T_\ell$ are equal, the conclusion of Theorem 1.3 is known to hold whenever the polynomials $n, n^2, \ldots, n^\ell$ are replaced by any family of linearly independent polynomials $p_1, \ldots, p_\ell$, each having zero constant term [17] (it is known that this independence assumption is necessary [7]). We conjecture that a similar result holds for any family of commuting, invertible measure preserving transformations $T_1, \ldots, T_\ell$. And in fact again, the assumption that the transformations $T_1, \ldots, T_\ell$ commute may be superfluous.

In most cases where the family of polynomials $p_1, \ldots, p_\ell$ is not pairwise independent, for example when $p_1(n) = \ldots = p_\ell(n) = n^2$, the methods of the present article do not suffice to study the limiting behavior of the averages (2).[3] It is in cases like this that working with some kind of "pleasant" extension (using terminology from [2]) or "magic" extension (using terminology form [22]) of the system may offer an essential advantage (this is indeed the case when all the polynomials are linear).

---

[3]There are particular (but rather exceptional) cases of non-pairwise independent families of polynomials, where the methods of the present paper can be easily modified and combined with the known "linear" results to prove convergence. One such example is when $p_1(n) = n, \ldots, p_{\ell-1}(n) = n$, and $p_\ell(n)$ is a polynomial with a sufficiently large degree (degree $> 2^\ell$ makes the problem accessible to the "simple" methods of the Appendix).

1.4. **General conventions and notation.** By a *system* we mean a Lebesgue probability space $(X, \mathcal{X}, \mu)$, endowed with a single, or several commuting, invertible measure preserving transformations, acting on X.

For notational convenience, all functions are implicitly assumed to be real valued, but straightforward modifications of our arguments, definitions, etc, can be given for complex valued functions as well.

We say that *the averages of the sequence* $(a_n)_{n \in \mathbb{N}}$ *converge* to some limit L, and we write

$$\lim_{N-M \to +\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} a_n = L,$$

if the averages of a_n on any sequence of intervals whose lengths tend to infinity converge to L. We use similar formulations for the lim sup and for limits in function spaces.

Lastly, the following notation will be used throughout the article: $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{T}^k = \mathbb{R}^k/\mathbb{Z}^k$, $Tf = f \circ T$, $e(t) = e^{2\pi i t}$.

2. Background in ergodic theory and nilmanifolds

2.1. **Background in ergodic theory.** Let $(X, \mathcal{X}, \mu, T)$ be a system.

Factors. A *homomorphism* from $(X, \mathcal{X}, \mu, T)$ to a system $(Y, \mathcal{Y}, \nu, S)$ is a measurable map $\pi \colon X' \to Y'$, where X' is a T-invariant subset of X and Y' is an S-invariant subset of Y, both of full measure, such that $\mu \circ \pi^{-1} = \nu$ and $S \circ \pi(x) = \pi \circ T(x)$ for $x \in X'$. When we have such a homomorphism we say that the system $(Y, \mathcal{Y}, \nu, S)$ is a *factor* of the system $(X, \mathcal{X}, \mu, T)$. If the factor map $\pi \colon X' \to Y'$ can be chosen to be bijective, then we say that the systems $(X, \mathcal{X}, \mu, T)$ and $(Y, \mathcal{Y}, \nu, S)$ are *isomorphic* (bijective maps on Lebesgue spaces have measurable inverses).

A factor can be characterized (modulo isomorphism) by $\pi^{-1}(\mathcal{Y})$, which is a T-invariant sub-σ-algebra of $\mathcal{B}$, and conversely any T-invariant sub-σ-algebra of $\mathcal{B}$ defines a factor. By a classical abuse of terminology we denote by the same letter the σ-algebra $\mathcal{Y}$ and its inverse image by π. In other words, if $(Y, \mathcal{Y}, \nu, S)$ is a factor of $(X, \mathcal{X}, \mu, T)$, we think of $\mathcal{Y}$ as a sub-σ-algebra of $\mathcal{X}$. A factor can also be characterized (modulo isomorphism) by a T-invariant subalgebra $\mathcal{F}$ of $L^\infty(X, \mathcal{X}, \mu)$, in which case $\mathcal{Y}$ is the sub-σ-algebra generated by $\mathcal{F}$, or equivalently, $L^2(X, \mathcal{Y}, \mu)$ is the closure of $\mathcal{F}$ in $L^2(X, \mathcal{X}, \mu)$.

Inverse limits. We say that $(X, \mathcal{X}, \mu, T)$ is an *inverse limit of a sequence of factors* $(X, \mathcal{X}_j, \mu, T)$ if $(\mathcal{X}_j)_{j \in \mathbb{N}}$ is an increasing sequence of T-invariant sub-σ-algebras such that $\bigvee_{j \in \mathbb{N}} \mathcal{X}_j = \mathcal{X}$ up to sets of measure zero.

Conditional expectation. If $\mathcal{Y}$ is a T-invariant sub-σ-algebra of $\mathcal{X}$ and $f \in L^1(\mu)$, we write $\mathbb{E}(f|\mathcal{Y})$, or $\mathbb{E}_\mu(f|\mathcal{Y})$ if needed, for the conditional expectation of f with respect to $\mathcal{Y}$. We will frequently make use of the identities

$$\int \mathbb{E}(f|\mathcal{Y}) \, d\mu = \int f \, d\mu \text{ and } T \, \mathbb{E}(f|\mathcal{Y}) = \mathbb{E}(Tf|\mathcal{Y}).$$

We say that a function f is orthogonal to $\mathcal{Y}$, and we write $f \perp \mathcal{Y}$, when it has a zero conditional expectation on $\mathcal{Y}$. If a function $f \in L^\infty(\mu)$ is measurable with respect to the factor $\mathcal{Y}$, we write $f \in L^\infty(\mathcal{Y}, \mu)$.

Ergodic decomposition. We write $\mathcal{I}$, or $\mathcal{I}(T)$ if needed, for the σ-algebra $\{A \in \mathcal{X} : T^{-1}A = A\}$ of invariant sets. A system is *ergodic* if all the T-invariant sets have measure either 0 or 1.

Let $x \mapsto \mu_x$ be a regular version of the conditional measures with respect to the σ-algebra $\mathcal{I}$. This means that the map $x \mapsto \mu_x$ is $\mathcal{I}$-measurable, and for very bounded measurable function f we have

$$\mathbb{E}_\mu(f|\mathcal{I})(x) = \int f \, d\mu_x \text{ for } \mu\text{-almost every } x \in X.$$

Then the *ergodic decomposition* of μ is

$$\mu = \int \mu_x \, d\mu(x).$$

The measures μ_x have the additional property that for μ-almost every $x \in X$ the system $(X, \mathcal{X}, \mu_x, T)$ is ergodic.

The rational Kronecker factor. For every $d \in \mathbb{N}$ we define $\mathcal{K}_d = \mathcal{I}(T^d)$. The rational Kronecker factor is

$$\mathcal{K}_{\mathrm{rat}} = \bigvee_{d \geq 1} \mathcal{K}_d.$$

We write $\mathcal{K}_{\mathrm{rat}}(T)$, or $\mathcal{K}_{\mathrm{rat}}(T, \mu)$, when needed. This factor is spanned by the family of functions

$$\{f \in L^\infty(\mu) : T^d f = f \text{ for some } d \in \mathbb{N}\},$$

or, equivalently, by the family

$$\{f \in L^\infty(\mu) : Tf = e(a) \cdot f \text{ for some } a \in \mathbb{Q}\}.$$

If $\mathbb{E}_\mu(f_1 | \mathcal{K}_{\mathrm{rat}}(T, \mu)) = 0$, then we have, for μ-almost every $x \in X$, that $\mathbb{E}_{\mu_x}(f_1 | \mathcal{K}_{\mathrm{rat}}(T, \mu_x)) = 0$ (see Lemma 3.2 in [17]).

2.2. The seminorms $\|\cdot\|_k$ and the factors $\mathcal{Z}_k$.

Sections 3 and 4 of [23] contain constructions that associate to every ergodic system a sequence of measures, seminorms, and factors. It is the case that for these constructions the hypothesis of ergodicity is not needed. Most properties remain valid, and can be proved in exactly the same manner, for general, not necessarily ergodic systems. We review the definitions and results we need in the sequel.

Let $(X, \mathcal{X}, \mu, T)$ be a system. We write $\mu = \int \mu_x \, d\mu(x)$ for the ergodic decomposition of μ.

Definition of the seminorms. For every $k \geq 1$, we define a measure $\mu^{[k]}$ on X^{2^k} invariant under $T \times T \times \cdots \times T$ (2^k times), by

$$\mu^{[1]} = \mu \times_{\mathcal{I}(T)} \mu = \int \mu_x \times \mu_x \, d\mu(x) \; ;$$

$$\text{for } k \geq 1, \; \mu^{[k+1]} = \mu^{[k]} \times_{\mathcal{I}(T \times T \times \cdots \times T)} \mu^{[k]}.$$

Writing $\underline{x} = (x_0, x_1, \ldots, x_{2^k - 1})$ for a point of X^{2^k}, we define a seminorm $\|\cdot\|_k$ on $L^\infty(\mu)$ by

$$\|f\|_k = \left(\int \prod_{i=0}^{2^k - 1} f(x_i) \, d\mu^{[k]}(\underline{x}) \right)^{1/2^k}.$$

That $\|\cdot\|_k$ is a seminorm can be proved as in [23], and also follows from the estimate (7) below. If needed, we are going to write $\|\cdot\|_{k,\mu}$, or $\|\cdot\|_{k,T}$.

By the inductive definition of the measures $\mu^{[k]}$ we have

$$\|f\|_1 = \|\mathbb{E}(f|\mathcal{I})\|_{L^2(\mu)} \ ; \tag{5}$$

$$\|f\|_{k+1}^{2^{k+1}} = \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \|f \cdot T^n f\|_k^{2^k}. \tag{6}$$

This can be considered as an alternate definition of the seminorms (assuming one first establishes existence of the limit in (6)).

For functions $f_0, f_1, \ldots, f_{2^k-1} \in L^\infty(\mu)$, the next inequality ([23], Lemma 3.9) follows from the definition of the measures by a repeated use of the Cauchy-Schwarz inequality

$$\left| \int \prod_{i=0}^{2^k-1} f_i(x_i) \, d\mu^{[k]}(\underline{x}) \right| \le \prod_{i=0}^{2^k-1} \|f_i\|_k. \tag{7}$$

Seminorms and ergodic decomposition. By induction, for every $k \in \mathbb{N}$ we have

$$\mu^{[k]} = \int (\mu_x)^{[k]} \, d\mu(x). \tag{8}$$

Therefore, for every function $f \in L^\infty(\mu)$ we have

$$\|f\|_{k,\mu}^{2^k} = \int \|f\|_{k,\mu_x}^{2^k} \, d\mu(x). \tag{9}$$

The factors $\mathcal{Z}_k$. For every $k \ge 1$, an invariant σ-algebra $\mathcal{Z}_k$ on X is constructed exactly as in Section 4 of [23]. It satisfies the same property as in Lemma 4.3 of [23]

$$\text{for } f \in L^\infty(\mu), \ \mathbb{E}_\mu(f|\mathcal{Z}_{k-1}) = 0 \text{ if and only if } \|f\|_{k,\mu} = 0. \tag{10}$$

Equivalently, one has

$$L^\infty(\mathcal{Z}_{k-1}, \mu) = \left\{ f \in L^\infty(\mu) : \int f \cdot g \, d\mu = 0 \text{ for every } g \in L^\infty(\mu) \text{ with } \|g\|_k = 0 \right\}. \tag{11}$$

In particular, if $f \in L^\infty(\mu)$ is measurable with respect to $\mathcal{Z}_{k-1}$ and satisfies $\|f\|_k = 0$, then $f = 0$. Therefore,

$$\| \cdot \|_k \text{ is a norm on } L^\infty(\mathcal{Z}_{k-1}, \mu).$$

If further clarification is needed, we are going to write $\mathcal{Z}_{k,\mu}$, or $\mathcal{Z}_{k,T}$. If $f \in L^\infty(\mu)$, then it follows from (9) and (10) that

$$\mathbb{E}_\mu(f|\mathcal{Z}_{k,\mu}) = 0 \text{ if and only if } \mathbb{E}_{\mu_x}(f|\mathcal{Z}_{k,\mu_x}) = 0 \text{ for } \mu\text{-almost every } x \in X. \tag{12}$$

Furthermore, if $f \in L^\infty(\mu)$, then

$$f \in L^\infty(\mathcal{Z}_{k,\mu}, \mu) \text{ if and only if } f \in L^\infty(\mathcal{Z}_{k,\mu_x}, \mu_x) \text{ for } \mu\text{-almost every } x \in X.$$

The first implication is non-trivial to establish though, due to various measurability problems. We prove this in Corollary 3.3 below.

For every $\ell \in \mathbb{N}$ one has $\|f\|_{1,T^\ell} \ll_\ell \|f\|_{2,T}$ (see proof of Proposition 2 in [24]). Using this and the inductive definition of the seminorms (6), one sees that $\|f\|_{k,T^\ell} \ll_{k,\ell} \|f\|_{k+1,T}$. Therefore,

$$\text{if } \ f \perp L^\infty(\mathcal{Z}_{k,T}, \mu) \ \text{ then } \ f \perp L^\infty(\mathcal{Z}_{k-1,T^\ell}, \mu) \ \text{ for every } \ell \in \mathbb{N}. \tag{13}$$

2.3. **Nilsystems, nilsequences, the structure of $\mathcal{Z}_k$.** A *nilmanifold* is a homogeneous space $X = G/\Gamma$ where G is a nilpotent Lie group, and Γ is a discrete cocompact subgroup of G. If $G_{k+1} = \{e\}$, where G_k denotes the k-the commutator subgroup of G, we say that X is a *k-step nilmanifold*.

A k-step nilpotent Lie group G acts on G/Γ by left translation where the translation by a fixed element $a \in G$ is given by $T_a(g\Gamma) = (ag)\Gamma$. By m_X we denote the unique probability measure on X that is invariant under the action of G by left translations (called the *normalized Haar measure*), and by $\mathcal{G}/\Gamma$ we denote the Borel σ-algebra of G/Γ. Fixing an element $a \in G$, we call the system $(G/\Gamma, \mathcal{G}/\Gamma, m_X, T_a)$ a *k-step nilsystem*.

If $X = G/\Gamma$ is a k-step nilmanifold, $a \in G$, $x \in X$, and $f \in \mathcal{C}(X)$, we call the sequence $(f(a^n x))_{n \in \mathbb{N}}$ a *basic k-step nilsequence*. A *k-step nilsequence*, is a uniform limit of *basic k-step nilsequences*. As is easily verified, the collection of k-step nilsequences, with the topology of uniform convergence, forms a closed algebra. We caution the reader that in other articles the term k-step nilsequence is used for what we call here basic k-step nilsequence, and in some instances the function f is assumed to satisfy weaker or stronger conditions than continuity.

The connection between the factors $\mathcal{Z}_k$ of a given ergodic system and nilsystems is given by the following structure theorem ([23], Lemma 4.3, Definition 4.10, and Theorem 10.1):

Theorem 2.1 ([23])**.** *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system and $k \in \mathbb{N}$.*
Then the system $(X, \mathcal{Z}_k, \mu, T)$ is a (measure theoretic) inverse limit of k-step nilsystems.

Remark. In fact, in [26] it is shown that, for ergodic systems, the factor $(X, \mathcal{Z}_k, \mu, T)$ is (measurably) isomorphic to a topological inverse limit of ergodic k-step nilsystems (for a definition see [26]). We are going to use this fact later.[4]

2.4. **Characteristic factors for linear averages.** Using successive applications of van der Corput's lemma, the following can be proved by induction on ℓ as in Theorem 12.1 of [23] (the $\ell = 2$ case follows for example from Theorem 2.1 in [21]):

Theorem 2.2. *Let $\ell \geq 2$, $(X, \mathcal{X}, \mu, T)$ be a system, $f_1, \ldots, f_\ell \in L^\infty(\mu)$, and $a_1, \ldots, a_\ell$ be distinct non-zero integers. Suppose that $f_i \perp \mathcal{Z}_{\ell-1}$ for some $i \in \{1, \ldots, \ell\}$.*
Then the averages

$$\frac{1}{N - M} \sum_{N=M}^{N-1} f_1(T^{a_1 n}x) \cdot \ldots \cdot f_\ell(T^{a_\ell n}x)$$

converge to 0 in $L^2(\mu)$.

3. A KEY APPROXIMATION PROPERTY

In this section we are going to prove the following key approximation result:

Proposition 3.1. *Let $(X, \mathcal{X}, \mu, T)$ be a system (not necessarily ergodic) and suppose that $f \in L^\infty(\mathcal{Z}_k, \mu)$ for some $k \in \mathbb{N}$.*
Then for every $\varepsilon > 0$ there exists a function $\tilde{f} \in L^\infty(\mu)$, with L^∞-norm bounded by $\|f\|_{L^\infty(\mu)}$, such that

---

[4]A *topological dynamical system* is a pair (Y, S) where Y is a compact metric space and $S\colon Y \to Y$ is a continuous transformation. If $(Y_i, S_i)_{i \in \mathbb{N}}$ is a sequence topological dynamical systems and $\pi_i\colon Y_{i+1} \to Y_i$ are factor maps, the *inverse limit* of the systems is defined to be the compact subset Y of $\prod_{i \in \mathbb{N}} Y_i$ given by $Y = \{(y_i)_{i \in \mathbb{N}}\colon \pi_i(y_{i+1}) = y_i\}$, with the induced infinite product metric and continuous transformation T.

(i) $\tilde{f} \in L^\infty(\mathcal{Z}_k, \mu)$ and $\left\| f - \tilde{f} \right\|_{L^1(\mu)} \leq \varepsilon$;

(ii) for μ-almost every $x \in X$, the sequence $(\tilde{f}(T^n x))_{n \in \mathbb{N}}$ is a k-step nilsequence.

By [28], if $(a_n)_{n \in \mathbb{N}}$ is a k-step nilsequence, and $p \in \mathbb{Z}[t]$ is a polynomial of degree $d \geq 1$, then the sequence $(a_{p(n)})_{n \in \mathbb{N}}$ is a (dk)-step nilsequence. Therefore, the function $\tilde{f}$ given by the proposition satisfies:

(iii) for μ-almost every $x \in X$ and every polynomial $p \in \mathbb{Z}[t]$ of degree $d \geq 1$, the sequence $\left(\tilde{f}(T^{p(n)} x) \right)_{n \in \mathbb{N}}$ is a (dk)-step nilsequence.

If the system $(X, \mathcal{X}, \mu, T)$ is ergodic, then one can deduce Proposition 3.1 from Theorem 2.1 in a straightforward way. It turns out to be much harder to prove this result in the non-ergodic case (and this strengthening is crucial for our later applications), due to a non-trivial measurable selection problem that one has to overcome. We give the proof in the following subsections.

3.1. **Dual functions.** In this subsection, $(X, \mathcal{X}, \mu, T)$ is a system, and the ergodic decomposition of μ is $\mu = \int \mu_x \, d\mu(x)$. We remind the reader that we work with real valued functions only.

We define a family of functions that will be used in the proof of Proposition 3.1 and gather some basic properties they satisfy.

When f is a bounded measurable function on X, for every $N \in \mathbb{N}$, we write

$$A_N(f) = \frac{1}{N^k} \sum_{1 \leq n_1, \ldots, n_k \leq N} \prod_{\substack{\epsilon \in \{0,1\}^k, \\ \epsilon \neq 00\cdots0}} f(T^{n_1 \epsilon_1 + \cdots + n_k \epsilon_k} x).$$

It is known by Theorem 1.2 in [23] that the averages $A_N(f)$ converge in $L^2(\mu)$ (in fact by [1] they converge pointwise but we do not need this strengthening), and we define

$$\mathcal{D}_k f = \lim_{N \to \infty} A_N(f)$$

where the limit is taken in $L^2(\mu)$. If needed, we write $\mathcal{D}_{k,\mu} f$. The function $\mathcal{D}_k f$ satisfies ([23], Theorem 13.1): For every $g \in L^\infty(\mu)$, we have

$$(14) \qquad \int g \cdot \mathcal{D}_k f \, d\mu = \int g(x_0) \prod_{i=1}^{2^k - 1} f(x_i) \, d\mu^{[k]}(\underline{x})$$

where $\underline{x} = (x_0, x_1, \ldots, x_{2^k - 1}) \in X^{2^k}$. In particular, by the definition of $\|f\|_k$, we have

$$(15) \qquad \int f \cdot \mathcal{D}_k f \, d\mu = \|f\|_k^{2^k},$$

and by inequality (7), for every function $g \in L^\infty(\mu)$ we have

$$(16) \qquad \left| \int g \cdot \mathcal{D}_k f \, d\mu \right| \leq \|g\|_k \cdot \|f\|_k^{2^k - 1}.$$

Example 1. We have

$$\mathcal{D}_1 f = \lim_{N \to \infty} \frac{1}{N} \sum_{n_1=1}^{N} T^n f = \mathbb{E}(f | \mathcal{I}).$$

Also,

$$\mathcal{D}_2 f(x) = \lim_{N \to \infty} \frac{1}{N^2} \sum_{1 \le n_1, n_2 \le N} f(T^{n_1} x) \cdot f(T^{n_2} x) \cdot f(T^{n_1 + n_2} x).$$

If T is an ergodic rotation on the circle $\mathbb{T}$ with the Haar measure $m_{\mathbb{T}}$, then an easy computation gives

$$(\mathcal{D}_2 f)(x) = \int_{\mathbb{T}} \int_{\mathbb{T}} f(x + s) \cdot f(x + t) \cdot f(x + s + t) \, dm_{\mathbb{T}}(s) \, dm_{\mathbb{T}}(t).$$

Notice that in this case the function $\mathcal{D}_2 f(x)$ may be non-constant, and no matter whether the function f is continuous or not, the function $\mathcal{D}_2 f(x)$ is continuous on $\mathbb{T}$.123

We gather some additional basic properties of dual functions.

Proposition 3.2. *Let $(X, \mathcal{X}, \mu, T)$ be a system.*
Then for every $f \in L^{\infty}(\mu)$ and $k \in \mathbb{N}$ the following hold:

(i) *For μ-almost every $x \in X$, we have $\mathcal{D}_{k,\mu} f = \mathcal{D}_{k,\mu_x} f$ as functions of $L^{\infty}(\mu_x)$.*
(ii) *The function $\mathcal{D}_k f$ is $\mathcal{Z}_{k-1}$-measurable, in fact $\mathcal{D}_k f = \mathcal{D}_k \tilde{f}$ where $\tilde{f} = \mathbb{E}(f | \mathcal{Z}_{k-1})$.*
(iii) *Linear combinations of functions $\mathcal{D}_k f$ with $f \in L^{\infty}(\mu)$ are dense in $L^1(\mathcal{Z}_{k-1}, \mu)$.*

Proof. We show (i). The averages $A_N(f)$ converge to $\mathcal{D}_{k,\mu} f$ in $L^2(\mu)$. Therefore, there exists a subsequence of $A_N(f)$ that converges to $\mathcal{D}_{k,\mu} f$ almost everywhere with respect to μ. As a consequence, for μ-almost every $x \in X$, this subsequence converges to $\mathcal{D}_{k,\mu} f$ almost everywhere with respect to μ_x. On the other hand, by the definition of $\mathcal{D}_{k,\mu_x} f$, for μ-almost every $x \in X$ this subsequence also converges to $\mathcal{D}_{k,\mu_x} f$ in $L^2(\mu_x)$. The result follows.

We show (ii). Since the operation $\mathcal{D}_k$ maps $L^{\infty}(\mathcal{Z}_{k-1}, \mu)$ to itself, it suffices to establish the second claim. Let $g \in L^{\infty}(\mu)$. Using (14) and expanding f as $\tilde{f} + (f - \tilde{f})$, we see that $\int g \cdot \mathcal{D}_k f \, d\mu$ is equal to $\int g \cdot \mathcal{D}_k \tilde{f} \, d\mu$, plus integrals of the form

$$\int g(x_0) \cdot \prod_{i=1}^{2^k - 1} f_i(x_i) \, d\mu^{[k]}(\underline{x}),$$

where each of the functions f_i is equal to either $\tilde{f}$ or to $f - \tilde{f}$, and at least one of the functions f_i is equal to $f - \tilde{f}$. Since $\mathbb{E}(f - \tilde{f} | \mathcal{Z}_{k-1}) = 0$, by (10) we have $\|f - \tilde{f}\|_k = 0$, and by inequality (7), all these integrals are equal to zero. This establishes that $\int g \cdot \mathcal{D}_k f \, d\mu$ is equal to $\int g \cdot \mathcal{D}_k \tilde{f} \, d\mu$, and the announced result follows.

We show (iii). By duality, it suffices to show that if $g \in L^{\infty}(\mathcal{Z}_{k-1}, \mu)$ satisfies $\int g \cdot \mathcal{D}_k f \, d\mu = 0$ for every $f \in L^{\infty}(\mu)$, then $g = 0$. Taking $f = g$ gives $\int g \cdot \mathcal{D}_k g \, d\mu = 0$, and using (15) we get $\|g\|_k = 0$. Since $\| \cdot \|_k$ is a norm in $L^{\infty}(\mathcal{Z}_{k-1}, \mu)$ we deduce that $g = 0$. This completes the proof. □

Corollary 3.3. *Let $(X, \mathcal{X}, \mu, T)$ be a system and $f \in L^{\infty}(\mathcal{Z}_{k,\mu}, \mu)$ for some $k \in \mathbb{N}$.*
Then, for μ-almost every $x \in X$, we have $f \in L^{\infty}(\mathcal{Z}_{k,\mu_x}, \mu_x)$.

Proof. By part (iii) of Propositition 3.2, there exists a sequence $(f_n)_{n \in \mathbb{N}}$, of finite linear combinations of functions of the form $\mathcal{D}_{k+1} \phi$ where $\phi \in L^{\infty}(\mu)$, such that $f_n \to f$ in $L^1(\mu)$. Passing to a subsequence, we can assume that $f_n \to f$ almost everywhere with respect to μ. As a consequence, for μ-almost every $x \in X$, we have $f_n \to f$ almost everywhere with respect to μ_x.

Furthermore, by parts (i) and (ii) of Propositition 3.2, we have that $f_n \in L^\infty(\mathscr{Z}_{k,\mu_x}, \mu_x)$ for μ-almost every x and every $n \in \mathbb{N}$. The announced result follows.   □

3.2. **Proof of Proposition 3.1.** In order to prove Proposition 3.1 we are going to make use of two ingredients. The first is a continuity property of dual functions (it follows from Proposition 5.2 and Lemma 5.8 in [26]):

Theorem 3.4 ([26]). *Suppose that the topological dynamical system (Y, S) is a topological inverse limit of minimal $(k-1)$-step nilsystems, and let m_Y be the unique S-invariant measure in Y.*

Then for every $f \in L^\infty(m_Y)$, and $k \in \mathbb{N}$, the function $\mathcal{D}_k f$ coincides m_Y-almost everywhere with a continuous function in Y.

Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system and let $(Y_k, \mathcal{Y}_k, m_k, S_k)$ be a topological inverse limit of minimal nilsystems that is measure theoretically isomorphic to the factor $(X, \mathscr{Z}_k, \mu, T)$ (see the remark following Theorem 2.1). With $\pi_k \colon X \to Y_k$ we denote the measure preserving isomorphism that identifies $(X, \mathscr{Z}_k, \mu, T)$ with $(Y_k, \mathcal{Y}_k, m_k, S_k)$. Using this notation we have:

Corollary 3.5. *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system, $f \in L^\infty(\mu)$, and $k \in \mathbb{N}$.*

Then there exists $g \in \mathcal{C}(Y_{k-1})$ such that $\mathcal{D}_k f$ coincides μ-almost everywhere with the function $g \circ \pi_{k-1}$.

Proof. By part (ii) of Proposition 3.2 we have that $\mathcal{D}_k f = \mathcal{D}_k \tilde{f}$ where $\tilde{f} = \mathbb{E}(f | \mathscr{Z}_{k-1})$. Therefore, we can assume that $f \in L^\infty(\mathscr{Z}_{k-1}, \mu)$. Writing $f = \phi \circ \pi_{k-1}$ for some $\phi \in L^\infty(m_{k-1})$, we have $\mathcal{D}_k f = (\mathcal{D}_k \phi) \circ \pi_{k-1}$. The announced result now follows from Theorem 3.4.   □

The second ingredient is Theorem 1.1 of [26], which gives a characterization of nilsequences that uses only local information about the sequence. To give here the exact statement would necessitate to introduce definitions and notation that we are not going to use in the sequel, so we choose to only state an immediate consequence that we need.

Theorem 3.6 ([26]). *Let $(a_s(n))_{n \in \mathbb{N}}$ be a collection of sequences indexed by a set S.*

Then for every $k \in \mathbb{N}$ the set of $s \in S$ for which the sequence $(a_s(n))_{n \in \mathbb{N}}$ is a k-step nilsequence belongs to the σ-algebra spanned by sets of the form $A_{l,m,n} = \{s \in S \colon |a_s(m) - a_s(n)| \leq 1/l\}$, where $l, m, n \in \mathbb{N}$.

Using this, we immediately deduce the following measurability property:

Corollary 3.7. *Let $(X, \mathcal{X}, \mu, T)$ be a system, $f \in L^\infty(\mu)$, and $k \in \mathbb{N}$.*

Then the set $A_f = \{x \in X \colon (f(T^n x))_{n \in \mathbb{N}}$ is a k-step nilsequence$\}$ is measurable.

We are now ready for the proof of Proposition 3.1.

Proof of Proposition 3.1. First notice that if a function $\tilde{f} \in L^\infty(\mu)$ satisfies properties (i) and (ii), then the function $g = \min(|\tilde{f}|, \|f\|_{L^\infty(\mu)}) \cdot \operatorname{sign}(\tilde{f})$ has L^∞-norm bounded by $\|f\|_{L^\infty(\mu)}$ and still satisfies properties (i) and (ii) (we used here that $\min(|a_n|, M) \cdot \operatorname{sign}(a_n)$ is a nilsequence if a_n is). So it suffices to find $\tilde{f} \in L^\infty(\mu)$ that satisfies properties (i) and (ii).

Since, by part (iii) of Proposition 3.2, for every $k \in \mathbb{N}$, linear combinations of functions of the form $\mathcal{D}_{k+1,\mu} \phi$ with $\phi \in L^\infty(\mu)$ are dense in $L^1(\mathscr{Z}_{k,\mu}, \mu)$, we can assume that f is of the

form $\mathcal{D}_{k+1,\mu}\phi$ for some $\phi \in L^\infty(\mu)$. Hence, it suffices to show that for every $\phi \in L^\infty(\mu)$ we have $\mu(A_{\mathcal{D}_{k+1,\mu}\phi}) = 1$, where

$$A_{\mathcal{D}_{k+1,\mu}\phi} = \{x \in X \colon ((\mathcal{D}_{k+1,\mu}\phi)(T^n x))_{n \in \mathbb{N}} \text{ is a } k\text{-step nilsequence}\}.$$

Let $\mu = \int \mu_x \, d\mu(x)$ be the ergodic decomposition of the measure μ. Since by Corollary 3.7 the set $A_{\mathcal{D}_{k+1,\mu}\phi}$ is μ-measurable, it suffices to show that $\mu_x(A_{\mathcal{D}_{k+1,\mu}\phi}) = 1$ for μ-almost every $x \in X$. By part (i) of Proposition 3.2 we have for μ-almost every $x \in X$ that $\mathcal{D}_{k+1,\mu}\phi = \mathcal{D}_{k+1,\mu_x}\phi$ as functions of $L^\infty(\mu_x)$. As a consequence, it remains to show that $\mu_x(A_{\mathcal{D}_{k+1,\mu_x}\phi}) = 1$ for μ-almost every $x \in X$.

We have therefore reduced matters to establishing that $\mu(A_{\mathcal{D}_{k+1,\mu}\phi}) = 1$ for ergodic systems and $\phi \in L^\infty(\mu)$. Using Corollary 3.5 and the notation introduced there, we get that there exists a function $g \in \mathcal{C}(Y_k)$ such that for μ-almost every $x \in X$ we have

$$(\mathcal{D}_{k+1,\mu}\phi)(x) = g(\pi_k x).$$

As a consequence, for μ-almost every $x \in X$, we have

$$(\mathcal{D}_{k+1,\mu}\phi)(T^n x) = g(S_k^n \pi_k x) \text{ for every } n \in \mathbb{N}.$$

Since (Y_k, S_k) is a topological inverse limit of nilsystems and $g \in \mathcal{C}(Y_k)$, for every $y \in Y_k$ the sequence $(g(S_k^n y))_{n \in \mathbb{N}}$ is a k-step nilsequence. We conclude that indeed $\mu(A_{\mathcal{D}_{k+1}\phi}) = 1$. This completes the proof. $\qquad \square$

4. A characteristic factor for the highest degree iterate: Two transformations

In this section and the next one, we are going to prove Theorem 1.2 under the additional assumption that the function corresponding to the highest degree polynomial iterate satisfies the stated orthogonality assumption. For example, if $\deg(p_1) > \deg(p_i)$ for $i = 2, \dots, \ell$, we assume that $f_1 \perp \mathcal{Z}_{k,T_1}$ for some $k \in \mathbb{N}$.

In fact our method necessitates that we prove a more general result (Proposition 5.1). This result is also going to be used in Section 6, when we deal with the polynomials of lower degree.

However, since the proof is notationally heavy, we present it first in the case of two commuting transformations. In the next section we give a sketch of the proof for the general case, focusing on the few points where the differences are significant.

In this section, we show:

Proposition 4.1. *Let $(X, \mathcal{X}, \mu, T_1, T_2)$ be a system and $f_1, \dots, f_m \in L^\infty(\mu)$. Let $(\mathcal{P}, \mathcal{Q})$ be a nice ordered family of pairs of polynomials, with degree d (all notions are defined in Section 4.2).*

Then there exists $k = k(d, m) \in \mathbb{N}$ such that: If $f_1 \perp \mathcal{Z}_{k,T_1}$, then the averages

$$(17) \qquad \frac{1}{N - M} \sum_{n=M}^{N-1} f_1(T_1^{p_1(n)} T_2^{q_1(n)} x) \cdot \ldots \cdot f_m(T_1^{p_m(n)} T_2^{q_m(n)} x)$$

converge to 0 in $L^2(\mu)$.

Applying this to the nice family $(\mathcal{P}, \mathcal{Q})$ where $\mathcal{P} = (p_1, 0)$ and $\mathcal{Q} = (0, p_2)$, we get:

Corollary 4.2. *Let $(X, \mathcal{X}, \mu, T_1, T_2)$ be a system and $f_1, f_2 \in L^\infty(\mu)$. Let p_1 and p_2 be integer polynomials with $d = \deg(p_1) > \deg(p_2)$.*

Then there exists $k = k(d)$ such that, if $f_1 \perp \mathcal{Z}_{k,T_1}$, then the averages

$$\frac{1}{N-M} \sum_{n=M}^{N-1} f_1(T_1^{p_1(n)} x) \cdot f_2(T_2^{p_2(n)} x)$$

converge to 0 in $L^2(\mu)$.

4.1. **A simple example.** We give here a very simple example in order to explain our strategy. In the appendix we consider slightly more general averages and get more precise results (the main drawback of these simpler arguments is that they do not allow us to treat any two polynomials with distinct degrees). Let $(X, \mathcal{X}, \mu, T_1, T_2)$ be a system and $f_1, f_2 \in L^\infty(\mu)$.

Claim. *If $f_1 \perp \mathcal{Z}_{2,T_1}$, then the averages*

$$(18) \qquad \frac{1}{N-M} \sum_{n=M}^{N-1} f_1(T_1^{n^2} x) \cdot f_2(T_2^n x)$$

converge to 0 in $L^2(\mu)$.

Using van der Corput's Lemma it suffices to show that for every $h_1 \in \mathbb{N}$, the averages in n of

$$\int f_1(T_1^{n^2} x) \cdot f_2(T_2^n x) \cdot f_1(T_1^{(n+h_1)^2} x) \cdot f_2(T_2^{n+h_1} x) \, d\mu(x)$$

converge to 0. After composing with T_2^{-n} and using the Cauchy-Schwarz inequality, we reduce matters to showing that the averages in n of

$$f_1(T_1^{n^2} T_2^{-n} x) \cdot f_1(T_1^{(n+h_1)^2} T_2^{-n} x)$$

converge to 0 in $L^2(\mu)$. Using van der Corput's Lemma one more time, we reduce matters to showing that for every fixed $h_1 \in \mathbb{N}$, for every large enough $h_2 \in \mathbb{N}$, the averages in n of

$$\int f_1(T_1^{n^2} T_2^{-n} x) \cdot f_1(T_1^{(n+h_1)^2} T_2^{-n} x) \cdot f_1(T_1^{(n+h_2)^2} T_2^{-n-h_2} x) \cdot f_1(T_1^{(n+h_1+h_2)^2} T_2^{-n-h_2} x) \, d\mu(x)$$

converge to 0, or equivalently, that the averages in n of

$$(19) \quad \int f_1(x) \cdot f_1(T_1^{2nh_1+h_1^2} x) \cdot f_1(T_1^{2nh_2+h_2^2} T_2^{-h_2} x) \cdot f_1(T_1^{2n(h_1+h_2)+(h_1+h_2)^2} T_2^{-h_2} x) \, d\mu(x)$$

converge to 0.

The important property of this last average is that it involves only constant iterates of the transformation T_2 (for h_1, h_2 fixed). Therefore, we can apply the known results about the convergence of averages of a single transformation. It follows from Theorem 2.2 that the averages in n of (19) converge to 0 for all $h_1, h_2 \in \mathbb{N}$ such that the linear polynomials $2h_1 n, 2h_2 n, 2(h_1 + h_2)n$ are distinct, that is, for all $h_1, h_2 \in \mathbb{N}$ with $h_1 \neq h_2$. The claim follows.

We will come back to this example in Section 6.1.

4.2. **Families of pairs and their type.** In this subsection we follow [8] with some changes on the notation, in order to define the type of a family of pairs of polynomials.

4.2.1. *Families of pairs of polynomials.* Let $m \in \mathbb{N}$. Given two ordered families of polynomials

$$\mathcal{P} = (p_1, \ldots, p_m), \quad \mathcal{Q} = (q_1, \ldots, q_m)$$

we define the *ordered family of pairs of polynomials* $(\mathcal{P}, \mathcal{Q})$ as follows

$$(\mathcal{P}, \mathcal{Q}) = \big((p_1, q_1), \ldots, (p_m, q_m)\big).$$

The reader is advised to think of this family as an efficient way to record the polynomial iterates that appear in (17).

The maximum of the degrees of the polynomials in the families $\mathcal{P}$ and $\mathcal{Q}$ is called the *degree of the family* $(\mathcal{P}, \mathcal{Q})$.

For convenience of exposition, if pairs of constant polynomials appear in $(\mathcal{P}, \mathcal{Q})$ we remove them, and henceforth we assume:

> All families $(\mathcal{P}, \mathcal{Q})$ that we consider do not contain pairs of constant polynomials.

4.2.2. *Definition of type.* We fix an integer $d \geq 1$ and restrict ourselves to families $(\mathcal{P}, \mathcal{Q})$ of degree $\leq d$.

We say that two polynomials $p, q \in \mathbb{Z}[t]$ are *equivalent*, and write $p \sim q$, if they have the same degree and the same leading coefficient. Equivalently, $p \sim q$ if and only if $\deg(p - q) < \min\{\deg(p), \deg(q)\}$

We define $\mathcal{Q}'$ to be the following set (possibly empty)

$$\mathcal{Q}' = \{q_i \in \mathcal{Q} : p_i \text{ is constant}\}.$$

For $i = 1, \ldots, d$, let $w_{1,i}, w_{2,i}$ be the number of distinct non-equivalent classes of polynomials of degree i in $\mathcal{P}$ and $\mathcal{Q}'$ correspondingly.

We define the *(matrix) type* of the family $(\mathcal{P}, \mathcal{Q})$ to be the $2 \times d$ matrix

$$\begin{pmatrix} w_{1,d} & \cdots & w_{1,1} \\ w_{2,d} & \cdots & w_{2,1} \end{pmatrix}.$$

If $\mathcal{Q}'$ is empty, then all the elements of the second row are taken to be 0. For example, with $d = 4$, the family

$$\big((n^2, n^4), \ (n^2 + n, n), \ (2n^2, 2n), \ (0, n^3), \ (0, n)\big)$$

has type

$$\begin{pmatrix} 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

We order the types lexicographically; we start by comparing the first element of the first row of each matrix, and after going through all the elements of the first row, we compare the elements of the second row of each matrix, and so on. In symbols: given two $2 \times d$ matrices $W = (w_{i,j})$ and $W' = (w'_{i,j})$, we say that $W > W'$ if: $w_{1,d} > w'_{1,d}$, or $w_{1,d} = w'_{1,d}$ and $w_{1,d-1} > w'_{1,d-1}$, ..., or $w_{1,i} = w'_{1,i}$ for $i = 1, \ldots, d$ and $w_{2,d} > w'_{2,d}$, and so on.

For example

$$\begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix} > \begin{pmatrix} 2 & 1 \\ \star & \star \end{pmatrix} > \begin{pmatrix} 2 & 0 \\ \star & \star \end{pmatrix} > \begin{pmatrix} 1 & \star \\ \star & \star \end{pmatrix} > \begin{pmatrix} 0 & \star \\ \star & \star \end{pmatrix} \geq \begin{pmatrix} 0 & 0 \\ \star & \star \end{pmatrix} \geq \begin{pmatrix} 0 & 0 \\ 0 & \star \end{pmatrix} \geq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

where in the place of the stars one can put any collection of non-negative integers.

An important observation is that although for a given type W there is an infinite number of possible types $W' < W$, we have

Lemma 4.3. *Every decreasing sequence of types of families of polynomial pairs is stationary.*

Therefore, if some operation reduces the type, then after a finite number of repetitions it is going to terminate. This is the basic principle behind all the PET induction arguments used in the literature and in this article.

4.3. **Nice families and the van der Corput operation.** In this subsection we define a class of families of pairs of polynomials that we are going to work with in the sequel, and an important operation that preserves such families and reduces their type.

4.3.1. *Nice families.* Let $\mathcal{P} = (p_1, \ldots, p_m)$ and $\mathcal{Q} = (q_1, \ldots, q_m)$.

Definition. We call the ordered family of pairs of polynomials $(\mathcal{P}, \mathcal{Q})$ *nice* if

(i) $\deg(p_1) \geq \deg(p_i)$ for $i = 1, \ldots, m$;

(ii) $\deg(p_1) > \deg(q_i)$ for $i = 1, \ldots, m$;

(iii) $\deg(p_1 - p_i) > \deg(q_1 - q_i)$ for $i = 2, \ldots, m$.

(Notice that a consequence of (iii) is that $p_1 - p_i \neq \text{const}$ for $i = 2, \ldots, m$.)

As an example, if a nice family consists of m pairs of polynomials and has degree 1, then we have: $\deg(p_1) = 1$, $\deg(p_i) \leq 1$, $\deg(q_i) = 0$ for $i = 1, \ldots, m$, and $\deg(p_1 - p_i) = 1$ for $i = 2, \ldots, m$. It follows that the type of this family is

$$\text{(20)} \qquad \begin{pmatrix} 0 & \cdots & 0 & k \\ 0 & \cdots & 0 & 0 \end{pmatrix}$$

for some $k \in \mathbb{N}$ with $k \leq m$.

4.3.2. *The van der Corput operation.* Given a family $\mathcal{P} = (p_1, \ldots, p_m)$, $p \in \mathbb{Z}[t]$, and $h \in \mathbb{N}$, we define

$$S_h \mathcal{P} = (p_1(n+h), \ldots, p_m(n+h)) \text{ and } \mathcal{P} - p = (p_1 - p, \ldots, p_m - p).$$

Given a family of pairs of polynomials $(\mathcal{P}, \mathcal{Q})$, a pair $(p, q) \in (\mathcal{P}, \mathcal{Q})$, and $h \in \mathbb{N}$, we define the following operation

$$(p, q, h)\text{-vdC}(\mathcal{P}, \mathcal{Q}) = (\tilde{P}, \tilde{Q})^*$$

where

$$\tilde{P} = (S_h \mathcal{P} - p, \mathcal{P} - p), \quad \tilde{Q} = (S_h \mathcal{Q} - q, \mathcal{Q} - q),$$

and * is the operation that removes all pairs of constant polynomials from a given family of pairs of polynomials. A more explicit form of the family $(p, q, h)\text{-vdC}(\mathcal{P}, \mathcal{Q})$ is

$$\big((S_h p_1 - p, S_h q_1 - q), \ldots, (S_h p_m - p, S_h q_m - q), (p_1 - p, q_1 - q), \ldots, (p_m - p, q_m - q)\big)^*.$$

Notice that if the family $(\mathcal{P}, \mathcal{Q})$ has degree d and contains m pairs of polynomials, then for every $h \in \mathbb{N}$, the family $(p, q, h)\text{-vdC}(\mathcal{P}, \mathcal{Q})$ has degree at most d and contains at most $2m$ pairs of polynomials.

4.4. **An example.** In order to explain our method we give an example that is somewhat more complicated than the example of Section 4.1. When we study the limiting behavior of the averages

$$\frac{1}{N-M} \sum_{n=M}^{N-1} f_1(T_1^{n^3} x) \cdot f_2(T_2^{n^2} x),$$

we define $\mathcal{P} = (n^3, 0)$, $\mathcal{Q} = (0, n^2)$, and introduce the family of pairs of polynomials

$$(\mathcal{P}, \mathcal{Q}) = \left((n^3, 0), (0, n^2) \right).$$

This family is nice and has type $\left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix} \right)$. Applying the vdC operation with $(p, q) = (0, n^2)$ and $h \in \mathbb{N}$, we arrive to the new family

$$(0, n^2, h)\text{-vdC}(\mathcal{P}, \mathcal{Q}) = (\tilde{P}_h, \tilde{Q}_h)$$

where

$$\tilde{P}_h = ((n+h)^3, 0, n^3, 0), \quad \tilde{Q}_h = (-n^2, 2hn + h^2, -n^2, 0) ;$$

then the corresponding family of pairs is

$$\left(\left((n+h)^3, -n^2 \right), \left(0, 2hn + h^2 \right), \left(n^3, -n^2 \right) \right).$$

The important point is that for every $h \in \mathbb{N}$ this new family is also nice and has smaller type, namely $\left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right)$. Translating back to ergodic theory, we get the averages

$$\frac{1}{N-M} \sum_{n=M}^{N-1} \tilde{f}(T_1^{(n+h)^3} T_2^{-n^2} x) \cdot \tilde{g}(T_2^{2hn+h^2} x) \cdot \tilde{h}(T_1^{n^3} T_2^{-n^2} x)$$

for some choice of functions $\tilde{f}, \tilde{g}, \tilde{h} \in L^\infty(\mu)$. Concerning the choice of these functions, the only important thing for our purposes is that $\tilde{f} = f_1$.

4.5. **The general strategy.** As was the case in the previous example, we are going to show that if we are given a nice family $(\mathcal{P}, \mathcal{Q})$ with $\deg(p_1) \geq 2$, then it is always possible to find appropriate $(p, q) \in (\mathcal{P}, \mathcal{Q})$ so that for all large enough $h \in \mathbb{N}$ the operation (p, q, h)-vdC leads to a nice family that has smaller type. Our objective is, after successively applying the operation (p, q, h)-vdC, to finally get nice families of degree 1, and thus with matrix type of the form (20).

Translating this back to ergodic theory, we get multiple ergodic averages (with certain parameters) where: (i) only linear iterates of the transformation T_1 appear and the iterates of T_2 are constant, and (ii) the "first" iterate of T_1 is applied to the "first" function of the original average. The advantage now is that the limiting behavior of such averages can be treated easily using the well developed theory of multiple ergodic averages involving a single transformation.

Let us remark though that in practice this process becomes cumbersome very quickly. For instance, in the example of Section 4.4, for every $h \in \mathbb{N}$, the next (p_h, q_h, h')-vdC operation uses $(p_h, q_h) = (0, 2hn + h^2)$ and leads to a family with matrix type $\left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix} \right)$. The subsequent vdC operation leads to a family with matrix type $\left(\begin{smallmatrix} 0 & 7 & 0 \\ 0 & 0 & 0 \end{smallmatrix} \right)$. One then has to apply the vdC operation a huge number of times (it is not even easy to estimate this number) in order to reduce the matrix type to the form (20). So even in the case of two commuting transformations, it is practically impossible to spell out the details of how this process works when both polynomial iterates are non-linear.

4.6. **Choosing a good vdC operation.** The next lemma is the key ingredient used to carry out the previous plan. To prove it we are going to use freely the following easy to prove fact: If p, q are two non-constant polynomials and $p \sim q$, then $\deg(p - q) \leq \deg(p) - 1$, and with the possible exception of one $h \in \mathbb{Z}$ we have $\deg(S_h p - q) = \deg(p) - 1$.

Lemma 4.4. *Let $(\mathcal{P}, \mathcal{Q})$ be a nice family of pairs of polynomials, and suppose that $\deg(p_1) \geq 2$. Then there exists $(p, q) \in (\mathcal{P}, \mathcal{Q})$, such that for every large enough $h \in \mathbb{N}$, the family (p, q, h)-vdC$(\mathcal{P}, \mathcal{Q})$ is nice and has strictly smaller type than that of $(\mathcal{P}, \mathcal{Q})$.*

Proof. Let $\mathcal{P} = (p_1, \ldots, p_m)$, $\mathcal{Q} = (q_1, \ldots, q_m)$, then for $(p, q) \in (\mathcal{P}, \mathcal{Q})$ and $h \in \mathbb{N}$ the family (p, q, h)-vdC$(\mathcal{P}, \mathcal{Q})$ is an ordered family of pairs of polynomials, all of them of the form

$$(S_h p_i - p, S_h q_i - q), \quad \text{or} \quad (p_i - p, q_i - q).$$

We choose (p, q) as follows: If $\mathcal{Q}'$ is non-empty, then we take $p = 0$ and let q to be a polynomial of smallest degree in $\mathcal{Q}'$. Then the first row of the matrix type remains unchanged, and the second row will get "reduced", leading to a smaller matrix type. Suppose now that $\mathcal{Q}'$ is empty. If $\mathcal{P}$ consists of a single polynomial p_1, then we choose $(p, q) = (p_1, q_1)$ and the result follows. Therefore, we can assume that $\mathcal{P}$ contains a polynomial other than p_1. We consider two cases. If $p \sim p_1$ for all $p \in \mathcal{P}$, then we choose $(p, q) = (p_1, q_1)$. Otherwise, we choose $(p, q) \in (\mathcal{P}, \mathcal{Q})$ such that $p \nsim p_1$ and p is a polynomial in $\mathcal{P}$ with minimal degree (such a choice exists since p_1 has the highest degree in $\mathcal{P}$).

In all cases, for every $h \in \mathbb{N}$, the first row of the matrix type of (p, q, h)-vdC$(\mathcal{P}, \mathcal{Q})$ is "smaller" than that of $(\mathcal{P}, \mathcal{Q})$, and as a consequence the new family has strictly smaller type.

It remains to verify that for every large enough $h \in \mathbb{N}$ the ordered family of pairs of polynomials (p, q, h)-vdC$(\mathcal{P}, \mathcal{Q})$ is nice. We remark that, by construction, the first polynomial pair in this family is $(S_h p_1 - p, S_h q_1 - q)$.

Claim. *Property* (i) *holds for every $h \in \mathbb{N}$.*

Equivalently, we claim that

$$\deg(S_h p_1 - p) \geq \max\{\deg(p_i - p), \deg(S_h p_i - p)\} \text{ for } i = 1, \ldots, m.$$

If $p \nsim p_1$, then $\deg(S_h p_1 - p) = \deg(p_1)$ and the claim follows from our assumption $\deg(p_1) \geq \deg(p_i)$ for $i = 1, \ldots, m$. If $p \sim p_1$, then by the choice of the polynomial p we have $p = p_1$ and $p \sim p_i$ for $i = 1, \ldots, m$. As a result, $\deg(S_h p_1 - p) = \deg(p_1) - 1$ and $\max\{\deg(p_i - p), \deg(S_h p_i - p)\} \leq \deg(p_1) - 1$, proving the claim.

Claim. *Property* (ii) *holds for every $h \in \mathbb{N}$.*

Equivalently, we claim that

$$\deg(S_h p_1 - p) > \max\{\deg(q_i - q), \deg(S_h q_i - q)\} \text{ for } i = 1, \ldots, m.$$

If $p \nsim p_1$, then $\deg(S_h p_1 - p) = \deg(p_1)$ and the claim follows since by assumption $\deg(p_1) > \deg(q_i)$ for $i = 1, \ldots, m$. If $p \sim p_1$, then by the choice of p we have $(p, q) = (p_1, q_1)$ and $p \sim p_i$ for $i = 1, \ldots, m$. By hypothesis we have

$$(21) \qquad \deg(q_i - q_1) < \deg(p_i - p_1) \leq \deg(p_1) - 1 = \deg(S_h p_1 - p_1).$$

It remains to verify that $\deg(S_h p_1 - p_1) > \deg(S_h q_i - q_1)$. To see this we express $S_h q_i - q_1$ as $(S_h q_i - q_i) + (q_i - q_1)$. If q_i is non-constant, then the first polynomial has degree $\deg(q_i) - 1 <$

$\deg(p_1)-1 = \deg(S_h p_1 - p_1)$. If q_i is constant, then it has degree $0 < \deg(p_1)-1 = \deg(S_h p_1 - p_1)$ (we used here that $\deg(p_1) \geq 2$). Furthermore, by (21) the second polynomial has degree $\deg(q_i - q_1) < \deg(S_h p_1 - p_1)$. This proves the claim.

Claim. *Property* (iii) *holds for all except finitely many values of h.*

Equivalently, we claim that
$$\deg(S_h p_1 - S_h p_i) > \deg(S_h q_1 - S_h q_i), \text{ for } i = 2, \ldots, m,$$
and
$$\deg(S_h p_1 - p_i) > \deg(S_h q_1 - q_i), \text{ for } i = 1, \ldots, m.$$
The first estimate follows immediately from our hypothesis $\deg(p_1 - p_i) > \deg(q_1 - q_i)$ for $i = 2, \ldots, m$. It remains to verify the second estimate. If $p_i \nsim p_1$, then $\deg(S_h p_1 - p_i) = \deg(p_1)$ and the claim follows since by hypothesis $\deg(p_1) > \deg(q_i)$ for $i = 1, \ldots, m$. Suppose now that $p_i \sim p_1$. Then $\deg(S_h p_1 - p_i) = \deg(p_1) - 1$, with the possible exception of one $h \in \mathbb{N}$ (hence we get at most $m-1$ exceptional values of h). So it remains to verify that $\deg(S_h q_1 - q_i) < \deg(p_1) - 1$. To see this we express $S_h q_1 - q_i$ as $(S_h q_1 - q_1) + (q_1 - q_i)$. The first polynomial has degree $\deg(q_1) - 1 < \deg(p_1) - 1$ if q_1 is non-constant, and degree $0 < \deg(p_1) - 1$ (we used that $\deg(p_1) \geq 2$) if q_1 is constant. The second polynomial has degree $\deg(q_1 - q_i) < \deg(p_1 - p_i) \leq \deg(p_1) - 1$ since $p_i \sim p_1$. This establishes the claim and completes the proof. $\square$

We say that a subset of $\mathbb{N}^k$ is *good* if it is of the form

(22)
$$\{h_1 \geq c_1, \ h_2 \geq c_2(h_1), \ldots, \ h_k \geq c_k(h_1, \ldots, h_{k-1})\}$$

for some $c_i \colon \mathbb{N}^{i-1} \to \mathbb{N}$. The next lemma will be used in order to prove that the level k of the characteristic factors $\mathcal{Z}_{k,T_i}$ considered in Theorem 1.2 depends only on the number and the maximum degree of the polynomials involved.

Lemma 4.5. *Let $(\mathcal{P}, \mathcal{Q})$ be a nice family with degree $d \geq 2$ that contains m pairs of polynomials. Suppose that we successively apply the (p, q, h)-vdC operation for appropriate choices of $p, q \in \mathbb{Z}[t]$ and $h \in \mathbb{N}$, as described in the previous lemma, each time getting a nice family of pairs of polynomials with strictly smaller matrix type.*

Then after a finite number of operations we get, for a good set of parameters, nice families of pairs of polynomials of degree 1. Moreover, the number of operations needed can be bounded by a function of d and m alone.

Remark. The exact dependency on d and m seems neither easy nor very useful to pin down; it appears to be a tower of exponentials the length of which depends on d and m.

Proof. We fix $d \geq 2$. The first statement follows immediately from Lemma 4.3.

We denote by $W(\mathcal{P}, \mathcal{Q})$ the matrix type of a given family $(\mathcal{P}, \mathcal{Q})$, and by $N(\mathcal{P}, \mathcal{Q})$ the number of operations mentioned in the statement needed to get the particular matrix type.

First we claim that it suffices to show the following: For every nice family $(\mathcal{P}, \mathcal{Q})$ with degree d, containing at most m polynomials, we have

(23)
$$N(\mathcal{P}, \mathcal{Q}) \leq f(W(\mathcal{P}, \mathcal{Q}), m)$$

for some function f, with the obvious domain, and range in the non-negative integers. Indeed, since there exists a finite number of possible matrix types for a family $(\mathcal{P}, \mathcal{Q})$ with degree at

most d and containing at most m polynomials (in fact there are at most $(m+1)^{2d}$ such matrix types), we have

$$N(\mathcal{P}, \mathcal{Q}) \le F(d, m) = \max_W (f(W, m))$$

where W ranges over all possible matrix types of nice families with degree d that contain at most m pairs of polynomials. This proves our claim.

Next we turn our attention to establishing (23). Let

$$W_1 = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Notice that W_1 is the smallest matrix type (with respect to the order introduced before) that can appear as a first coordinate entry in the domain of f. We define f recursively as follows:

(24) $\qquad f(W_1, m) = 0$ for every $m \in \mathbb{N}$, and $\quad f(W, m) = \max_{W' < W} (f(W', 2m)) + 1$

where the maximum is taken over the finitely many possible matrix types W' of nice families of degree at most d that contain at most $2m$ pairs of polynomials.

Since every (p, q, h)-vdC operation of the previous lemma preserves nice families of pairs of polynomials, does not increase their degree, reduces their matrix type, and at most doubles the number m of (non-constant) pairs of polynomials in the family, a straightforward induction on the type $W(\mathcal{P}, \mathcal{Q})$ establishes (23) with f defined by (24). This completes the proof. $\qquad \square$

4.7. **Proof of Proposition 4.1.** Let $(\mathcal{P}, \mathcal{Q})$ be a nice family of pairs of polynomials where $\mathcal{P} = (p_1, \ldots, p_m)$ and $\mathcal{Q} = (q_1, \ldots, q_m)$ and let d be the degree of this family. We remind the reader that our goal is to show that there exists $k = k(d, m) \in \mathbb{N}$ such that: If $f_1 \perp \mathcal{Z}_{k, T_1}$, then the averages of

(25) $\qquad f_1(T_1^{p_1(n)} T_2^{q_1(n)} x) \cdot \ldots \cdot f_m(T_1^{p_m(n)} T_2^{q_m(n)} x)$

converge to 0 in $L^2(\mu)$.

(a) Suppose first that $\deg(p_1) = 1$. Since the family $(\mathcal{P}, \mathcal{Q})$ is nice, we have $\deg(p_i) = 1$ for $i = 1, \ldots, m$, all the polynomials $q_1, \ldots, q_m$ are constant, and $p_1 - p_i \neq$ const for $i = 1, \ldots, m$. In other words we are reduced to studying the limiting behavior of the averages in n of

$$f_1(T_1^{a_1 n + b_1} T_2^{c_1} x) \cdot f_2(T_1^{a_2 n + b_2} T_2^{c_2} x) \cdot \ldots \cdot f_m(T_1^{a_m n + b_m} T_2^{c_m} x)$$

where $a_i, b_i, c_i \in \mathbb{Z}$, $a_i \neq 0$, for $i = 1, \ldots, m$, and $a_1 \neq a_i$ for $i = 2, \ldots, m$. Suppose that $f_1 \perp \mathcal{Z}_{m-1, T_1}$, then also $T_2 f_1 \perp \mathcal{Z}_{m-1, T_1}$ (since T_1 and T_2 commute). By Theorem 2.2 the previous averages converge to 0 in $L^2(\mu)$, and as a consequence the same holds for the averages of (25).

(b) Suppose now that $\deg(p_1) \ge 2$. Our objective is to repeatedly use van der Corput's Lemma in order to reduce matters to the previously established linear case.

To begin with, using van der Corput's Lemma we see that in order to establish convergence to 0 for the averages of (25), it suffices to show that, for every sufficiently large $h \in \mathbb{N}$, the averages in n of

$$\int f_1(T_1^{p_1(n+h)} T_2^{q_1(n+h)} x) \cdot \ldots \cdot f_m(T_1^{p_m(n+h)} T_2^{q_m(n+h)} x) \cdot$$

$$f_1(T_1^{p_1(n)} T_2^{q_1(n)} x) \cdot \ldots \cdot f_m(T_1^{p_m(n)} T_2^{q_m(n)} x) \; d\mu$$

converge to 0. We compose with $T_1^{-p(n)}T_2^{-q(n)}$, where $(p,q) \in (\mathcal{P}, \mathcal{Q})$ is chosen as in Lemma 4.4, and use the Cauchy-Schwarz inequality. This reduces matters to showing that, for every sufficiently large $h \in \mathbb{N}$, the averages in n of

$$(26) \quad f_1(T_1^{p_1(n+h)-p(n)}T_2^{q_1(n+h)-q(n)}x) \cdot \ldots \cdot f_m(T_1^{p_m(n+h)-p(n)}T_2^{q_m(n+h)-q(n)}x) \cdot$$
$$f_1(T_1^{p_1(n)-p(n)}T_2^{q_1(n)-q(n)}x) \cdot \ldots \cdot f_m(T_1^{p_m(n)-p(n)}T_2^{q_m(n)-q(n)}x)$$

converge to 0 in $L^2(\mu)$. We remove the functions that happen to be composed with constant iterates of T and S, since they do not affect convergence to 0. This corresponds to the operation * defined in Section 4.3.2. We get multiple ergodic averages that correspond to the families of polynomials (p, q, h)-vdC$(\mathcal{P}, \mathcal{Q})$; our goal is to show convergence to 0 in $L^2(\mu)$ for every large enough $h \in \mathbb{N}$.

By Lemma 4.4, for every large enough $h \in \mathbb{N}$, the family (p, q, h)-vdC$(\mathcal{P}, \mathcal{Q})$ is nice, and its first pair is $(p_1(n+h) - p(n), q_1(n+h) - q(n))$. Notice also that, in (26) the iterate $T_1^{p_1(n+h)-p(n)}T_2^{q_1(n+h)-q(n)}$ is applied to the function f_1. We consider two cases depending on the degree of the polynomial $p_1(n+h) - p(n)$.

($\mathbf{b_1}$) If $\deg(p_1(n+h) - p(n)) = 1$, then we are reduced to the case (a) studied before. As we explained, if $f_1 \perp \mathcal{Z}_{2m,T_1}$, then the averages (26) converge to 0 in $L^2(\mu)$ for every large enough $h \in \mathbb{N}$. As a consequence, the averages (25) converge to 0 in $L^2(\mu)$.

($\mathbf{b_2}$) If $\deg(p_1(n+h) - p(n)) \geq 2$, then we can iterate the "van der Corput operation". By Lemma 4.5, there exists $k = k(d, m) \in \mathbb{N}$, such that after at most k such operations, we arrive to averages involving, for a good set of parameters G of the form (22), nice families of pairs of polynomials of the type studied in part (a). More precisely, we are left with studying the averages in n of

$$(27) \qquad g_1(T_1^{a_1 n + b_1}T_2^{c_1}x) \cdot \ldots \cdot g_{\tilde{m}}(T_1^{a_{\tilde{m}} n + b_{\tilde{m}}}T_2^{c_{\tilde{m}}}x)$$

where the functions g_i, and the integers a_i, b_i, c_i, depend on k parameters, and satisfy: (i) $g_1 = f_1$ (this last condition follows easily by the definition of the vdC-operation), and (ii) $a_1 \neq a_i$ for $i \in \{2, \ldots, \tilde{m}\}$. Our goal is to show convergence to 0 in $L^2(\mu)$ for the averages of (27) for this good set of parameters G. Then repeated uses of van der Corput's Lemma show that the averages of (25) converge to 0 in $L^2(\mu)$.

We proceed to establish our goal. Since the number of functions involved at most doubles after each vdC-operation is performed, we have $\tilde{m} \leq 2^k m$. It follows by Theorem 2.2 and properties (i) and (ii) above, that if $f_1 \perp \mathcal{Z}_{2^k m, T_1}$, then for every choice of parameters in the "good" set G, the averages of (27) converge to 0 in $L^2(\mu)$, establishing our goal. As a consequence, the averages of (25) converge to 0 in $L^2(\mu)$.

Concluding, if $f_1 \perp \mathcal{Z}_{2^k m, T_1}$, then in all cases we showed that the averages of (25) converge to 0 in $L^2(\mu)$. This completes the proof of Proposition 4.1. $\qquad \square$

5. A characteristic factor for the highest degree iterate: The general case

The next proposition is the generalization of Proposition 4.1 to the case of an arbitrary number of transformations. Its proof is very similar to the proof of Proposition 4.1 that was given in the previous section. To avoid unnecessary repetition, we define the concepts needed

in the proof of Proposition 5.1, and then only summarize its proof providing details only when non-trivial modifications of the arguments used in the previous section are needed.

Proposition 5.1. *Let $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ be a system, and $f_1, \ldots, f_m \in L^\infty(\mu)$. Suppose that $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ is a nice ordered family of ℓ-tuples of polynomials with degree d (all notions are defined below).*

Then there exists $k = k(d, \ell, m) \in \mathbb{N}$ such that: If $f_1 \perp \mathcal{Z}_{k,T_1}$, then the averages

$$\frac{1}{N-M} \sum_{n=M}^{N-1} f_1(T_1^{p_{1,1}(n)} \cdots T_\ell^{p_{\ell,1}(n)} x) \cdot \ldots \cdot f_m(T_1^{p_{1,m}(n)} \cdots T_\ell^{p_{\ell,m}(n)} x)$$

converge to 0 in $L^2(\mu)$.

Applying this result to the family $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ where $\mathcal{P}_1 = (p_1, 0, \ldots, 0)$, $\mathcal{P}_2 = (0, p_2, \ldots, 0)$, $\ldots$, $\mathcal{P}_\ell = (0, \ldots, 0, p_\ell)$, we get:

Corollary 5.2. *Let $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ be a system, and $f_1, \ldots, f_\ell \in L^\infty(\mu)$. Let $p_1, \ldots, p_\ell$ be integers polynomials with distinct degrees and highest degree $d = \deg(p_1)$.*

Then there exists $k = k(d, \ell)$ such that: If $f_1 \perp \mathcal{Z}_{k,T_1}$, then the averages

$$\frac{1}{N-M} \sum_{n=M}^{N-1} f_1(T_1^{p_1(n)} x) \cdot \ldots \cdot f_\ell(T_\ell^{p_\ell(n)} x)$$

converge to 0 in $L^2(\mu)$.

5.1. **Families of ℓ-tuples and their types.** In this subsection we follow [8] with some changes in the notation.

5.1.1. *Families of ℓ-tuples of polynomials.* Let $\ell, m \in \mathbb{N}$. Given ℓ ordered families of polynomials

$$\mathcal{P}_1 = (p_{1,1}, \ldots, p_{1,m}), \ldots, \mathcal{P}_\ell = (p_{\ell,1}, \ldots, p_{\ell,m})$$

we define an *ordered family of m polynomial ℓ-tuples* as follows

$$(\mathcal{P}_1, \ldots, \mathcal{P}_\ell) = \big((p_{1,1}, \ldots, p_{\ell,1}), \ldots, (p_{1,m}, \ldots, p_{\ell,m})\big).$$

The reader is advised to think of this family as an efficient way to record the polynomial iterates that appear in the average of

$$f_1(T_1^{p_{1,1}(n)} \cdots T_\ell^{p_{\ell,1}(n)} x) \cdot \ldots \cdot f_m(T_1^{p_{1,m}(n)} \cdots T_\ell^{p_{\ell,m}(n)} x).$$

The maximum of the degrees of the polynomials in the families $\mathcal{P}_1, \ldots, \mathcal{P}_\ell$ is called *the degree of the family* $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$.

For convenience of exposition, if ℓ-tuples of constant polynomials appear in $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ we remove them, and henceforth we assume:

All families $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ that we consider do not contain ℓ-tuples of constant polynomials.

5.1.2. *Definition of type.* We fix $d \geq 1$ and restrict ourselves to families of degree $\leq d$.

For $i = 1, \ldots, \ell$, we define $\mathcal{P}'_i$ to be the following set (possibly empty)

$$\mathcal{P}'_i = \{\text{non-constant } p_{i,j} \in \mathcal{P}_i \colon p_{i',j} \text{ is constant for } i' < i\}.$$

(It follows that $\mathcal{P}'_1$ is the set of non-constant polynomials is $\mathcal{P}_1$.)

For $i = 1, \ldots, \ell$ and $j = 1, \ldots, d$, we let $w_{i,j}$ be the number of distinct non-equivalent classes of polynomials of degree j in the family $\mathcal{P}'_i$.

We define the *(matrix) type* of the family $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ to be the matrix

$$\begin{pmatrix} w_{1,d} & \cdots & w_{1,1} \\ w_{2,d} & \cdots & w_{2,1} \\ \vdots & \cdots & \vdots \\ w_{\ell,d} & \cdots & w_{\ell,1} \end{pmatrix}.$$

For example, let $d = 4$, and consider the family of triples of polynomials

$$\big((n^2, n^4, n^4),\ (n^2 + n, 3n^3, 0),\ (2n^2, 0, 2n),\ (n, 2n, 0),$$
$$(0, n^3, n^4),\ (0, 2n^3, n^2),\ (0, 0, n^3),\ (0, 0, n^3 + 1)\big).$$

Since

$$\mathcal{P}'_1 = \{n^2, n^2 + n, 2n^2, n\}, \quad \mathcal{P}'_2 = \{n^3, 2n^3\}, \quad \mathcal{P}'_3 = \{n^3, n^3 + 1\},$$

the type of this family is

$$\begin{pmatrix} 0 & 0 & 2 & 1 \\ 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

As in Section 4.2.2, we order these types lexicographically: Given two $\ell \times d$ matrices $W = (w_{i,j})$ and $W' = (w'_{i,j})$, we say that the first is bigger than the second, and write $W > W'$, if $w_{1,d} > w'_{1,d}$, or $w_{1,d} = w'_{1,d}$ and $w_{1,d-1} > w'_{1,d-1}$, ..., or $w_{1,i} = w'_{1,i}$ for $i = 1, \ldots, d$ and $w_{2,d} > w'_{2,d}$, and so on. As for the types of families of pairs, we have:

Lemma 5.3. *Every decreasing sequence of types of families of polynomial ℓ-tuples is stationary.*

5.2. Nice families and the van der Corput operation. In this subsection we define a class of families of ℓ-tuples of polynomial that we are going to work with in the sequel, and an important operation that preserves such families and reduces their type.

5.2.1. *Nice families.* Let $\mathcal{P}_1 = (p_{1,1}, \ldots, p_{1,m})$, ..., $\mathcal{P}_1 = (p_{\ell,1}, \ldots, p_{\ell,m})$.

Definition. We call the ordered family of polynomial ℓ-tuples $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ *nice* if

   (i) $\deg(p_{1,1}) \geq \deg(p_{1,j})$ for $j = 1, \ldots, m$;

   (ii) $\deg(p_{1,1}) > \deg(p_{i,j})$ for $i = 2, \ldots, \ell,\ j = 1, \ldots, m$;

   (iii) $\deg(p_{1,1} - p_{1,j}) > \deg(p_{i,1} - p_{i,j})$ for $i = 2, \ldots, \ell,\ j = 2, \ldots, m$.

(Notice that a consequence of (iii) is that $p_{1,1} - p_{1,j}$ is not constant for $j = 2, \ldots, m$.)

The type of a nice family of degree 1 has only one non-zero entry, namely $w_{1,1}$.

5.2.2. *The van der Corput operation.* Given a family $\mathcal{P} = (p_1, \ldots, p_m)$, $p \in \mathbb{Z}[t]$, and $h \in \mathbb{N}$, we define $S_h\mathcal{P}$ and $\mathcal{P} - p$ as in Section 4.3.2. Given a family of ℓ-tuples of polynomials $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$, $(p_1, \ldots, p_\ell) \in (\mathcal{P}_1, \ldots \mathcal{P}_\ell)$, and $h \in \mathbb{N}$, we define the following operation

$$(p_1, \ldots, p_\ell, h)\text{-vdC}(\mathcal{P}_1, \ldots, \mathcal{P}_\ell) = (\tilde{P}_{1,h}, \ldots \tilde{P}_{\ell,h})^*$$

where

$$\tilde{P}_{i,h} = (S_h\mathcal{P}_i - p_i, \mathcal{P}_i - p_i).$$

for $i = 1, \ldots, \ell$, and * is the operation that removes all constant ℓ-tuples polynomials from a given family of ℓ-tuples polynomials. Notice that if $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ is a degree d family containing m polynomial ℓ-tuples, then for every $h \in \mathbb{N}$, the family $(p_1, \ldots, p_\ell, h)\text{-vdC}(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ has degree at most d and contains at most $2m$ polynomial ℓ-tuples.

5.3. **Choosing a good vdC operation.** As in the case of two transformations, our objective is starting with a nice family $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ to successively apply appropriate operations $(p_1, \ldots, p_\ell, h)\text{-vdC}(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ in order to arrive to nice families of polynomial ℓ-tuples with types that have only non-zero entry the entry $w_{1,1}$. This case then can be treated easily using known results that involve a single transformation.

Lemma 5.4. *Let $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ be a nice family with $\deg(p_{1,1}) \geq 2$.*
Then there exists $(p_1, \ldots, p_\ell) \in (\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ such that for every large enough $h \in \mathbb{N}$ the family $(p_1, \ldots, p_\ell, h)\text{-vdC}(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ is nice and has strictly smaller type than $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$.

Proof. We remind the reader that we have $\mathcal{P}_i = (p_{i,1}, \ldots, p_{i,m})$ for $i = 1, \ldots, \ell$. For $(p_1, \ldots, p_\ell) \in (\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$, the family $(p_1, \ldots, p_\ell, h)\text{-vdC}(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ consists of vectors of polynomials that have the form

$$(S_h p_{1,j} - p_1, \ldots, S_h q_{\ell,j} - p_\ell), \ j = 1, \ldots, m, \ \text{ or } \ (p_{1,j} - p_1, \ldots, p_{\ell,j} - p_\ell).$$

We choose $(p_1, \ldots, p_\ell)$ as follows:

If $\mathcal{P}'_\ell$ is non-empty, then we take $p_1 = \cdots = p_{\ell-1} = 0$ and p_ℓ to be a polynomial of smallest degree in $\mathcal{P}'_\ell$. Then for every $h \in \mathbb{N}$, the first $\ell - 1$ rows of the type will remain unchanged, and the last row will get "reduced", leading to a smaller matrix type. Similarly, if the families $\mathcal{P}'_\ell, \mathcal{P}'_{\ell-1}, \ldots, \mathcal{P}'_{i-1}$ are empty, and P'_i is non-empty for some $2 \leq i \leq \ell + 1$, then we take $p_1 = \cdots = p_{i-1} = 0$ and p_i to be a polynomial of smallest degree in $\mathcal{P}'_i$. Then for every $h \in \mathbb{N}$, the first $i - 1$ rows of the matrix type remain unchanged, and the i-the row will get "reduced", leading to a smaller matrix type.

Suppose now that the families $\mathcal{P}'_\ell, \mathcal{P}'_{\ell-1}, \ldots, \mathcal{P}'_2$ are empty. If $\mathcal{P}_1$ consists of a single polynomial, namely $p_{1,1}$, then we choose $(p_1, \ldots, p_\ell) = (p_{1,1}, \ldots, p_{\ell,1})$ and the result follows. Therefore, we can assume that $\mathcal{P}_1$ contains some polynomial other than $p_{1,1}$. We consider two cases. If $p \sim p_{1,1}$ for all $p \in \mathcal{P}_1$, then we choose $(p_1, \ldots, p_\ell) = (p_{1,1}, \ldots, p_{\ell,1})$. Otherwise, we choose $(p_1, \ldots, p_\ell) \in (\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ with $p_1 \nsim p_{1,1}$, and p_1 is a polynomial in $\mathcal{P}_1$ with smallest degree (such a choice exists since $p_{1,1}$ has the highest degree in $\mathcal{P}_1$). In all cases, for every $h \in \mathbb{N}$, the first row of the matrix type of $(p_1, \ldots, p_\ell, h)\text{-vdC}(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ is "smaller" than that of $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$.

It remains to verify that for every large enough $h \in \mathbb{N}$ the family $(p_1, \ldots, p_\ell, h)\text{-vdC}(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ is nice. This part is identical with the one used in Lemma 4.4 and so we omit it. $\qquad\square$

The proof of the next lemma is completely analogous to the proof of Lemma 4.5 in the previous section and so we omit it.

Lemma 5.5. *Let* $(\mathcal{P}_1, \ldots, \mathcal{P}_\ell)$ *be a nice family with degree* $d \geq 2$ *that contains* m *polynomial* ℓ*-tuples. Suppose that we successively apply the* $(p_1, \ldots, p_\ell, h)$*-vdC operation for appropriate choices of* $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$ *and* $h \in \mathbb{N}$*, as described in the previous lemma, each time getting a nice family of* ℓ*-tuples of polynomials with strictly smaller type.*

Then after a finite number of operations we get, for a good set of parameters, nice families of ℓ*-tuples of polynomials of degree* 1*. Moreover, the number of operations needed can be bounded by a function of* d*,* ℓ*, and* m*.*

5.4. **Proof of Proposition 5.1.** Using Lemma 5.4 and Lemma 5.5, the rest of the proof of Proposition 5.1 is completely analogous to the end of the proof of Proposition 4.1 given in Section 4.7 and so we omit it.

6. Characteristic factors for the lower degree iterates and proof of convergence

In this section we prove Theorem 1.2 and then Theorem 1.1.

6.1. **A simple example.** In order to explain our method, we continue with the example of Section 4.1, studying the limiting behavior of the averages of

$$(28) \qquad f_1(T_1^{n^2}x) \cdot f_2(T_2^n x).$$

We have shown that these averages converge to 0 in $L^2(\mu)$ whenever $f_1 \perp \mathcal{Z}_{2,T_1}$. We are therefore reduced to study these averages under the additional hypothesis that f_1 is measurable with respect to $\mathcal{Z}_{2,T_1}$.

Using the approximation property of Proposition 3.1, we further reduce matters to the case where, for μ-almost every $x \in X$, the sequence $(f_1(T^n x))_{n \in \mathbb{N}}$ is a 2-step nilsequence. Therefore, the sequence $(f_1(T^{n^2}x))_{n \in \mathbb{N}}$ is a 4-step nilsequence. We are left with studying the limiting behavior of the averages of

$$u_n(x) \cdot f_2(T_2^n x),$$

where $(u_n)_{n \in \mathbb{N}}$ is a uniformly bounded sequence of μ-measurable functions, such that $(u_n(x))_{n \in \mathbb{N}}$ is a 4-step nilsequence for μ-almost every $x \in X$.

In this particular case, Corollary 6.3 below suffices to show that the averages converge to 0 in $L^2(\mu)$ whenever $f_2 \perp \mathcal{Z}_{4,T_2}$. (For more intricate averages we need more elaborate results about weighted multiple averages.)

We are reduced to the case where f_1 is measurable with respect to $\mathcal{Z}_{2,T_1}$ and f_2 is measurable with respect to $\mathcal{Z}_{4,T_2}$. Applying Proposition 3.1 to these two functions, we reduce matters to the case where, for μ-almost every $x \in X$, the sequences $(f_1(T_1^{n^2}x))_{n \in \mathbb{N}}$ and $(f_2(T_2^n x))_{n \in \mathbb{N}}$ are finite step nilsequences. Therefore, for μ-almost every $x \in X$, the sequence (28) is a nilsequence and as a consequence its averages converge.

We introduce now the tools that we need to carry out the previous plan in our more general setup.

6.2. **Uniformity seminorms.** We follow [25]. Let $k \in \mathbb{N}$ be an integer. Let $(a_n)_{n \in \mathbb{Z}}$ be a bounded sequence of real numbers and $\mathbf{I} = (I_N)_{N \in \mathbb{N}}$ be a sequence of intervals whose lengths $|I_N|$ tend to infinity. We say that this sequence of intervals is k-*adapted* to the sequence (a_n), if for every $\underline{h} = (h_1, \ldots, h_k) \in \mathbb{N}^k$, the limit

$$c_{\underline{h}}(\mathbf{I}, a) := \lim_{N \to \infty} \frac{1}{|I_N|} \sum_{n \in I_N} \prod_{\epsilon \in \{0,1\}^k} a_{n + h_1 \epsilon_1 + \cdots + h_k \epsilon_k}$$

exists[5]. Clearly, every sequence of intervals whose lengths tend to infinity admits a subsequence which is adapted to the sequence (a_n).

Suppose that $\mathbf{I} = (I_N)_{N \in \mathbb{N}}$ is k-adapted to $(a_n)_{n \in \mathbb{Z}}$. We define

$$\|a\|_{\mathbf{I},k} := \Big(\lim_{H \to +\infty} \frac{1}{H^k} \sum_{1 \le h_1, \ldots, h_k \le H} c_{\underline{h}}(\mathbf{I}, a) \Big)^{1/2^k}.$$

Indeed, by Proposition 2.2 of [25], the above limit exists and is non-negative.

Lemma 6.1. *Let $(X, \mathcal{X}, \mu, T)$ be a system, $f \in L^\infty(\mu)$, and $\mathbf{I} = (I_N)_{N \in \mathbb{N}}$ be a sequence of intervals whose lengths tend to infinity. Suppose that $f \perp \mathcal{Z}_{k-1,\mu}$ for some $k \ge 2$.*

Then the sequence $\mathbf{I}$ admits a subsequence $\mathbf{I}' = (I_N')_{N \in \mathbb{N}}$ such that, for μ-almost every $x \in X$, $\mathbf{I}'$ is k-adapted to the sequence $(f(T^n x))_{n \in \mathbb{N}}$ and $\|f(T^n x)\|_{\mathbf{I}',k} = 0$.

Proof. Let $\mu = \int \mu_x \, d\mu(x)$ be the ergodic decomposition of μ. For $x \in X$, we write $a(x) = (a_n(x))_{n \in \mathbb{N}}$ for the sequence defined by $a_n(x) = f(T^n x)$.

By the Ergodic Theorem, for every $\underline{h} = (h_1, \ldots, h_k) \in \mathbb{N}^k$, the averages

$$\frac{1}{|I_N|} \sum_{n \in I_N} \prod_{\epsilon \in \{0,1\}^k} a_{n + h_1 \epsilon_1 + \cdots + h_k \epsilon_k}(x)$$

converge in $L^2(\mu)$. As a consequence, a subsequence of this sequence of averages converges μ-almost everywhere. This subsequence depends on the parameter $\underline{h}$, but since there are only countably many such parameters, by a diagonal argument we can find a subsequence $\mathbf{I}' = (I_N')_{N \in \mathbb{N}}$ such that for μ almost every $x \in X$ the limit

$$(29) \qquad c_{\underline{h}}(\mathbf{I}', a(x)) = \lim_{N \to +\infty} \frac{1}{|I_N'|} \sum_{n \in I_N'} \prod_{\epsilon \in \{0,1\}^k} a_{n + h_1 \epsilon_1 + \cdots + h_k \epsilon_k}(x)$$

exists for every choice of $\underline{h} = (h_1, \ldots, h_k) \in \mathbb{N}^k$. This means that, for μ-almost every $x \in X$, the sequence of intervals $\mathbf{I}'$ is k-adapted to the sequence $(a_n(x))_{n \in \mathbb{N}}$.

Furthermore, by the Ergodic Theorem, for every $\underline{h} \in \mathbb{N}^k$ the averages on the right hand side of (29) converge in $L^2(\mu)$ to

$$\mathbb{E}_\mu \Big(\prod_{\epsilon \in \{0,1\}^k} T^{h_1 \epsilon_1 + \cdots + h_k \epsilon_k} f \Big| \mathcal{I}(T) \Big)(x) = \int \prod_{\epsilon \in \{0,1\}^k} T^{h_1 \epsilon_1 + \cdots + h_k \epsilon_k} f \, d\mu_x.$$

Therefore, for μ-almost every $x \in X$, we have

$$c_{\underline{h}}(\mathbf{I}', a(x)) = \int \prod_{\epsilon \in \{0,1\}^k} T^{h_1 \epsilon_1 + \cdots + h_k \epsilon_k} f \, d\mu_x.$$

---

[5]In [25] it is assumed that the limit exists for $\underline{h} \in \mathbb{Z}^k$ but this does not change anything in the proofs.

Taking the average in $\underline{h}$, using the definition of $\mathcal{D}_k f$ (Section 3.1), and (15), we get for μ-almost every $x \in X$ that

$$\|a(x)\|_{\mathbf{I}',k} = \|f\|_{k,\mu_x}.$$

Since by hypothesis $\mathbb{E}_\mu(f|\mathcal{Z}_{k-1}) = 0$, by (10) we have $\|f\|_{k,\mu} = 0$, and as a consequence $\|f\|_{k,\mu_x} = 0$ for μ-almost every $x \in X$ by (9). This completes the proof. $\square$

We are also going to use the following result:

Theorem 6.2 ([25], Corollary 2.14)**.** *Let $(a_n)_{n\in\mathbb{N}}$ be a bounded sequence of real numbers, and $\mathbf{I} = (I_N)_{N\in\mathbb{N}}$ be a sequence of intervals that is k-adapted to this sequence for some $k \geq 2$. Suppose that $\|a_n\|_{\mathbf{I},k} = 0$.*
Then for every bounded $(k-1)$-step nilsequence u_n we have

$$\lim_{N\to\infty} \frac{1}{|I_N|} \sum_{n\in I_N} a_n u_n = 0.$$

Combining the results of this section, we can now prove:

Corollary 6.3. *Let $(X, \mathcal{X}, \mu, T)$ be a system and $f \in L^\infty(\mu)$. Let $(u_n(x))_{n\in\mathbb{N}}$ be a uniformly bounded sequence of μ-measurable functions such that, for μ-almost every $x \in X$, the sequence $(u_n(x))_{n\in\mathbb{N}}$ is a k-step nilsequence for some $k \geq 1$. Suppose that $f \perp \mathcal{Z}_{k,T}$.*
Then the averages

$$\frac{1}{N-M} \sum_{n=M}^{N-1} f(T^n x) \cdot u_n(x)$$

converge to 0 in $L^2(\mu)$.

Proof. It suffices to prove that every sequence of intervals $\mathbf{I} = (I_N)_{n\in\mathbb{N}}$ whose lengths tend to infinity admits a subsequence $\mathbf{I}' = (I'_N)_{n\in\mathbb{N}}$ such that

$$(30) \qquad \frac{1}{|I'_N|} \sum_{n\in I'_N} f(T^n x) \cdot u_n(x) \to 0 \text{ in } L^2(\mu).$$

Let $\mathbf{I}'$ be given by Lemma 6.1 (with k in place of $k-1$). For μ-almost every $x \in X$ we have $\|(f(T^n x))_{n\in\mathbb{N}}\|_{\mathbf{I}',k+1} = 0$. Theorem 6.2 gives that the averages in (30) converge to 0 pointwise and the asserted convergence to 0 in $L^2(\mu)$ follows from the bounded convergence theorem. This completes the proof. $\square$

6.3. Some weighted averages.

We are going to prove Theorem 1.2 by induction on the number of transformations involved. The next result is going to help us carry out the induction step.

Proposition 6.4. *Let $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ be a system and $f_1, \ldots, f_\ell \in L^\infty(\mu)$. Let $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$ be polynomials with distinct degrees and highest degree $d = \deg(p_1)$. Let $(u_n(x))_{n\in\mathbb{N}}$ be a uniformly bounded sequence of μ-measurable functions such that, for μ-almost every $x \in X$, the sequence $(u_n(x))_{n\in\mathbb{N}}$ is an s-step nilsequence for some $s \geq 1$.*
Then there exists $k = k(d, \ell, s) \in \mathbb{N}$ such that: If $f_1 \perp \mathcal{Z}_{k,T_1}$, then the averages

$$\frac{1}{N-M} \sum_{n=M}^{N-1} f_1(T_1^{p_1(n)} x) \cdot \ldots \cdot f_\ell(T_\ell^{p_\ell(n)} x) \cdot u_n(x)$$

converge to 0 in $L^2(\mu)$.

Proof. First suppose that $\deg(p_1) = 1$. Then the polynomials $p_2, \ldots, p_\ell$ are all constant. The polynomial p_1 has the form $p_1(n) = an + b$ for some integers a, b with $a \neq 0$. Applying Corollary 6.3 for $T_1^b f_1$ in place of f and T_1^a in place of T, and using (13) we get the announced result with $k = s + 1$.

Therefore, we can assume that $\deg(p_1) \geq 2$. The strategy of the proof is the same as in Corollary 6.3, but instead of the Ergodic Theorem used in the proof of Lemma 6.1, we use Proposition 5.1.

We assume that $f_1 \perp \mathcal{Z}_{k,T_1}$, where k is the integer $k(d, \ell, 2^s\ell)$ given by Proposition 5.1. In order to prove the announced convergence to 0, it suffices to show that every sequence of intervals $\mathbf{I} = (I_N)_{N \in \mathbb{N}}$ admits a subsequence $\mathbf{I}' = (I'_N)_{N \in \mathbb{N}}$ such that

$$(31) \qquad \frac{1}{|I'_N|} \sum_{n \in I'_N} f_1(T_1^{p_1(n)} x) \cdot \ldots \cdot f_\ell(T_\ell^{p_\ell(n)} x) \cdot u_n(x) \quad \text{converges to } 0 \text{ in } L^2(\mu).$$

We let $m = 2^s$, and for $x \in X$, let $a(x) = (a_n(x))_{n \in \mathbb{Z}}$ be the sequence given by

$$a_n(x) = f_1(T_1^{p_1(n)} x) \cdot \ldots \cdot f_\ell(T_\ell^{p_\ell(n)} x).$$

For $r_1, \ldots, r_m \in \mathbb{Z}$, we study the averages

$$\frac{1}{|I_N|} \sum_{n \in I_N} a_{n+r_1}(x) \ldots a_{n+r_m}(x).$$

Consider the following ℓ ordered families of polynomials, each consisting of ℓm polynomials:

$$\mathcal{P}_1 = \big(p_1(n + r_1), \ldots, p_1(n + r_m), 0, \ldots, 0, \; \ldots\ldots\ldots, 0, \ldots, 0\big)$$
$$\mathcal{P}_2 = \big(0, \ldots, 0, p_2(n + r_1), \ldots, p_2(n + r_m), \; \ldots\ldots\ldots, 0, \ldots, 0\big)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\mathcal{P}_\ell = \big(0, \ldots, 0, 0, \ldots, 0, \; \ldots\ldots\ldots, p_\ell(n + r_1), \ldots, p_\ell(n + r_m)\big)$$

Using that $\deg(p_1) \geq 2$ and $\deg(p_i) < \deg(p_1)$ for $i = 2, \ldots, \ell$, it is easy to check that this family is nice except if $r_1 \in \{r_2, \ldots, r_m\}$.

Using Proposition 5.1 (with $k = k(d, \ell, 2^s\ell)$) we have that the averages

$$\frac{1}{|I_N|} \sum_{n \in I_N} a_{n+r_1}(x) \cdot \ldots \cdot a_{n+r_m}(x)$$

converge to 0 in $L^2(\mu)$ for every $r_1, \ldots, r_m \in \mathbb{Z}$ with $r_1 \notin \{r_2, \ldots, r_m\}$. As in the proof of Lemma 6.1, there exists a subsequence $\mathbf{I}' = (I'_N)_{N \in \mathbb{N}}$ of the sequence of intervals $\mathbf{I}$ such that

$$\frac{1}{|I'_N|} \sum_{n \in I'_N} a_{n+r_1}(x) \cdot \ldots \cdot a_{n+r_m}(x) \to 0 \quad \mu\text{-almost everywhere}$$

for all choices of $r_1, \ldots, r_m \in \mathbb{Z}$ with $r_1 \notin \{r_2, \ldots, r_m\}$.

In particular, for every $h_1, \ldots, h_s \in \mathbb{N}$, we have

$$\frac{1}{|I'_N|} \sum_{n \in I'_N} \prod_{\epsilon \in \{0,1\}^s} a_{n + \epsilon_1 h_1 + \cdots + \epsilon_s h_s}(x) \to 0 \quad \mu\text{-almost everywhere}.$$

To see this, apply the previous convergence property when $\{r_1, \ldots, r_m\}$ is equal to the set $\{\epsilon_1 h_1 + \cdots + \epsilon_s h_s, \epsilon_i \in \{0, 1\}\}$ and $r_1 = 0$.

As a consequence, for μ-almost every $x \in X$, the sequence $\mathbf{I}'$ of intervals is s-adapted to the sequence $a(x)$, and $c_{\underline{h}}(\mathbf{I}', a(x)) = 0$ for every $\underline{h} \in \mathbb{N}^s$. Therefore, $\|a(x)\|_{\mathbf{I}',s} = 0$ for μ-almost every $x \in X$. By Theorem 6.2, we have

$$\frac{1}{|I_N'|} \sum_{n \in I_N'} a_n(x) \cdot u_n(x) \to 0 \quad \mu\text{-almost everywhere}$$

and (31) is proved. This completes the proof. $\square$

6.4. **Proof of Theorem 1.2.** We are now ready to prove Theorem 1.2. It is a special case (take u_n to be constant) of the following result:

Theorem 6.5. *Let* $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ *be a system and* $f_1, \ldots, f_\ell \in L^\infty(\mu)$. *Let* $p_1, \ldots, p_\ell$ *be polynomials with distinct degrees and maximum degree* d. *Let* $(u_n(x))_{n \in \mathbb{N}}$ *be a uniformly bounded sequence of measurable functions on* X *such that, for* μ-*almost every* $x \in X$, *the sequence* $(u_n(x))_{n \in \mathbb{N}}$ *is an* s-*step nilsequence.*

Then there exists $k = k(d, \ell, s)$ *with the following property: If* $f_i \perp \mathcal{Z}_{k,T_i}$ *for some* $i \in \{1, \ldots, \ell\}$, *then the averages*

$$(32) \qquad \frac{1}{N - M} \sum_{n=M}^{N-1} f_1(T^{p_1(n)} x) \cdot \ldots \cdot f_\ell(T^{p_\ell(n)} x) \cdot u_n(x)$$

converge to 0 *in* $L^2(\mu)$.

Proof. The proof goes by induction on the number ℓ of transformations. For $\ell = 1$, the result is the case $\ell = 1$ of Proposition 6.4. We take $\ell \geq 2$, assume that the results holds for $\ell - 1$ transformations, and we are going to prove that it holds for ℓ transformations.

Without loss of generality we can assume that $\deg(p_1) = d > \deg(p_i)$ for $2 \leq i \leq \ell$. By Proposition 6.4, there exists $k_0 = k_0(d, \ell, s)$ such that, if $f_1 \perp \mathcal{Z}_{k_0,T_1}$, then the averages (32) converge to 0 in $L^2(\mu)$. Therefore we can restrict ourselves to the case where

the function f_1 is measurable with respect to $\mathcal{Z}_{k_0,T_1}$.

By Proposition 3.1, for every $\varepsilon > 0$, there exists $\tilde{f}_1 \in L^\infty(\mu)$, measurable with respect to $\mathcal{Z}_{k_0,T_1}$, with $\left\| f_1 - \tilde{f}_1 \right\|_{L^2(\mu)} < \varepsilon$, and such that $(\tilde{f}_1(T_1^n x))_{n \in \mathbb{N}}$ is a k_0-step nilsequence for μ-almost every $x \in X$. By density, it suffices to prove the result under the additional hypothesis that

$(f_1(T_1^n x))_{n \in \mathbb{N}}$ *is a k_0-step nilsequence for μ-almost every $x \in X$.*

Then for μ-almost every $x \in X$, the sequence $(f_1(T_1^{p_1(n)} x))_{n \in \mathbb{N}}$ is a (dk_0)-step nilsequence. The sequence $(f_1(T_1^{p_1(n)} x) \cdot u_n(x))_{n \in \mathbb{N}}$ is the product of two k-step nilsequences where $k = \max(dk_0, s)$ and thus it is a k-step nilsequence. Therefore, the announced result follows from the induction hypothesis. This completes the induction and the proof. $\square$

6.5. **Proof of Theorem 1.1.** Let $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ be a system and $f_1, \ldots, f_\ell \in L^\infty(\mu)$. We assume that the polynomials $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$ have distinct degrees and we want to show that the averages

$$(33) \qquad \frac{1}{N - M} \sum_{n=M}^{N-1} f_1(T_1^{p_1(n)} x) \cdot \ldots \cdot f_\ell(T_\ell^{p_\ell(n)} x)$$

converge in $L^2(\mu)$.

By Theorem 1.2, there exists $k \in \mathbb{N}$ such that the averages (33) converge to 0 whenever $f_i \perp \mathcal{Z}_{k,T_i}$ for some $i \in \{1, \ldots, \ell\}$. Therefore, we can assume that for $i = 1, \ldots, \ell$, the function f_i is measurable with respect to $\mathcal{Z}_{k,T_i}$.

By Proposition 3.1, for every $\varepsilon > 0$, and for $i = 1, \ldots, \ell$, there exists a function $\tilde{f}_i \in L^\infty(\mu)$, measurable with respect to $\mathcal{Z}_{k,T_i}$, with $\left\| f_i - \tilde{f}_i \right\|_{L^2(\mu)} < \varepsilon$, and such that $(\tilde{f}_i(T^n x))_{n \in \mathbb{N}}$ is a k-step nilsequence for μ-almost every $x \in X$.

By density we can therefore assume that, for $i = 1, \ldots, \ell$, and for μ-almost every $x \in X$, $(f_i(T^n x))_{n \in \mathbb{N}}$ is a k-step nilsequence and as a consequence $(f_i(T^{p_i(n)} x))_{n \in \mathbb{N}}$ is a (dk)-step nilsequence. Then for μ-almost every $x \in X$, the average (33) is an average of a (dk)-step nilsequence, and therefore it converges by [28]. This completes the proof. $\square$

7. Lower bounds for powers

In this section we are going to prove Theorem 1.3.

We remark that a consequence of Theorem 1.1 is that all the limits of multiple ergodic averages mentioned in this section exist (in $L^2(\mu)$). As a result, we are allowed to write $\lim_{N-M \to \infty}$, where $\limsup_{N-M \to \infty}$ should have been used.

We start with some background material.

7.1. **Equidistribution properties on nilmanifolds.** We summarize some notions and results that will be needed later.

Polynomial sequences. Let G be a nilpotent Lie group. Let $X = G/\Gamma$ be a nilmanifold, where Γ is a discrete cocompact subgroup of G. Recall that for $a \in G$ we write $T_a \colon X \to X$ for the *translation* $x \mapsto ax$.

If $a_1, \ldots, a_\ell \in G$, and $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$, then a sequence of the form $g(n) = a_1^{p_1(n)} a_2^{p_2(n)} \cdots a_\ell^{p_\ell(n)}$ is called a *polynomial sequence in G*. If $x \in X$ and $(g(n))_{n \in \mathbb{N}}$ is a polynomial sequence in G, then the sequence $(g(n)x)_{n \in \mathbb{N}}$ is called a *polynomial sequence in X*.

Sub-nilmanifolds. If H is a closed subgroup of G and $x \in X$, then Hx may not be a closed subset of X (for example, take $X = \mathbb{R}/\mathbb{Z}$, $x = \mathbb{Z}$, and $H = \{k\sqrt{2} \colon k \in \mathbb{Z}\}$), but if it is closed, then the compact set Hx can be given the structure of a nilmanifold ([28]). More precisely, if $x = g\Gamma$, then Hx is closed if and only if $\Delta = H \cap g\Gamma g^{-1}$ is cocompact in H. In this case $Hx \simeq H/\Delta$, and $h \mapsto hg\Gamma$ induces the isomorphism from H/Δ onto Hx. We call any such set Hx a *sub-nilmanifold* of X.

Equidistribution. We say that the sequence $(g(n)x)_{n\in\mathbb{N}}$, with values in a nilmanifold X, is *equidistributed* (or *well distributed*) in a sub-nilmanifold Y of X, if for every $F \in \mathcal{C}(X)$ we have

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} F(g(n)x) = \int F \, dm_Y$$

where m_Y denotes the normalized Haar measure on Y.

For typographical reasons, we use the following notation:

Notation. If E is a subset of X, we denote by $\mathrm{cl}_X(E)$ the closure of E in X.

A fact that we are going to use repeatedly is that polynomial sequences are equidistributed in their orbit closure. More precisely:

Theorem 7.1 ([28])**.** *Let $X = G/\Gamma$ be a nilmanifold, $(g(n)x)_{n\in\mathbb{N}}$ be a polynomial sequence in X, and $Y = \mathrm{cl}_X\{g(n)x, n \in \mathbb{N}\}$.*

(i) *There exists $r \in \mathbb{N}$ such that the sequence $(g(rn)x)_{n\in\mathbb{N}}$ is equidistributed on some connected component of Y.*

(ii) *If Y is connected, then Y is a sub-nilmanifold of X, and for every $r \in \mathbb{N}$ the sequence $(g(rn)x)_{n\in\mathbb{N}}$ is equidistributed on Y.*

Ergodic elements. An element $a \in G$ is *ergodic*, or *acts ergodically on X*, if the sequence $(a^n\Gamma)_{n\in\mathbb{N}}$ is dense in X.

Suppose that $a \in G$ acts ergodically on X. Then for every $x \in X$ the sequence $(a^n x)_{n\in\mathbb{N}}$ is equidistributed in X. If X is assumed to be connected, then for every $r \in \mathbb{N}$ the element a^r also acts ergodically on X (this follows from part (iii) of Theorem 7.1). For general nilmanifolds X we can easily deduce the following result (with X_0 we denote the connected component of the element Γ): There exists $r_0 \in \mathbb{N}$ such that the nilmanifold X is the disjoint union of the sub-nilmanifolds $X_i = a^i X_0$, $i = 1, \ldots, r_0$, and a^r acts ergodically on each X_i for every $r \in r_0\mathbb{N}$.

The affine torus. If $X = G/\Gamma$ is a connected nilmanifold, the *affine torus* of X is defined to be the homogeneous space $A = G/([G_0, G_0]\Gamma)$, where by G_0 we denote the connected component of the identity element in G. The homogeneous space A can be smoothly identified in a natural way with the nilmanifold $G_0/([G_0, G_0](\Gamma \cap G_0))$, which is a finite dimensional torus, say $\mathbb{T}^m$ for some $m \in \mathbb{N}$. It is known ([15]) that, under this identification, G acts on A by unipotent affine transformations. This means that every $T_g: \mathbb{T}^m \to \mathbb{T}^m$ has the form $Tx = Sx + b$, for some unipotent homomorphism S of $\mathbb{T}^m$ and $b \in \mathbb{T}^m$.

Equidistribution criterion. If $X = G/\Gamma$ is a nilmanifold, then X is connected if and only if $G = G_0\Gamma$. In the sequel we need to establish some equidistribution properties of polynomial sequences on nilmanifolds. The next criterion is going to simplify our task:

Theorem 7.2 ([28])**.** *Let $X = G/\Gamma$ be a connected nilmanifold, $(g(n))_{n\in\mathbb{N}}$ be a polynomial sequence in G, and $x \in X$. Let $A = G/([G_0, G_0]\Gamma)$ be the affine torus of X and $\pi_A: X \to A$ be the natural projection.*

Then the sequence $(g(n)x)_{n\in\mathbb{N}}$ is equidistributed in X if and only if the sequence $(g(n)\pi_A(x))_{n\in\mathbb{N}}$ is equidistributed in A.

7.2. **An example.** In order to explain the strategy of the proof of Theorem 1.3 we use an example. Our goal is to show that for a given system $(X, \mathcal{X}, \mu, T_1, T_2)$, and set $A \in \mathcal{X}$, for every $\varepsilon > 0$, we have

$$\mu(A \cap T_1^{-n} A \cap T_2^{-n^2} A) \geq \mu(A)^3 - \varepsilon$$

for a set of $n \in \mathbb{N}$ that has bounded gaps.

After some manipulations that are explained in Section 7.3, we are left with showing that if $f_1 \perp \mathcal{K}_{\mathrm{rat}}(T_1)$ or $f_2 \perp \mathcal{K}_{\mathrm{rat}}(T_2)$, then the averages of

(34) $$f_1(T_1^n x) \cdot f_2(T_2^{n^2} x)$$

converge to 0 in $L^2(\mu)$. In fact, we are only going to be able to prove a somewhat more technical variation of this property (see Proposition 7.3), but the exact details are not important at this point.

By Theorem 1.2 we can assume that the function f_1 is $\mathcal{Z}_{k,T_1}$-measurable and the function f_2 is $\mathcal{Z}_{k,T_2}$-measurable for some $k \in \mathbb{N}$. For convenience, we also assume that the transformation T_1 is totally ergodic (meaning T_1^r is ergodic for every $r \in \mathbb{N}$). In this case, using Theorem 2.1 and an approximation argument, we can further reduce matters to the case where X is a connected nilmanifold, $\mu = m_X$, and $T_1 = T_a$ is an ergodic translation on X. The assumption that X is connected is important, and is a consequence of our simplifying assumption that the transformation T_1 is totally ergodic. Also, by Proposition 3.1, we can assume that for m_X-almost every $x \in X$ the sequence $u_n(x) = f_2(T_2^n x)$ is a finite step nilsequence.

After doing all these maneuvers our new goal becomes to establish the following result:

(a) Let X be a connected nilmanifold, a be an ergodic translation of X, and $\int f_1 \, dm_X = 0$. Let $(u_n)_{n \in \mathbb{N}}$ be a uniformly bounded sequence of measurable functions such that $(u_n(x))_{n \in \mathbb{N}}$ is a nilsequence for m_X-almost every $x \in X$. Then the averages of

$$f_1(a^n x) \cdot u_{n^2}(x)$$

converge to 0 in $L^2(m_X)$. (The conclusion fails if X is not connected.)

It is easy to see that (a) follows from the following result:

(a)' Let X be a connected nilmanifold and a be an ergodic translation of X. Let Y be a nilmanifold and b be an ergodic translation of Y. Then for m_X-almost every $x \in X$ we have: for every nilmanifold Y, every ergodic translation b of Y, and every $y \in Y$, the sequence

$$(a^n x, b^{n^2} y)$$

is equidistributed on the nilmanifold $X \times Y$.

We prove a variation of this result that suffices for our purposes in Lemma 7.6. This is the heart of our argument, and we prove it by (i) showing that it suffices to verify the announced equidistribution property when each translation a and b is given by an ergodic unipotent affine transformation on some finite dimensional torus, and then (ii) verify the announced equidistribution property for affine transformations using direct computations (see Lemma 7.5). It is in this second step that we make crucial use of the special structure of our polynomial iterates; our argument does not quite work for some other distinct degree polynomials iterates like n and $n^2 + n$. The key observation is that since all the coordinates of the sequence $(a^n x)$ have non-trivial linear part, and those of $(b^{n^2} y)$ have trivial linear part, for typical values of $x \in X$, it is impossible for the coordinates of the sequences $(a^n x)$ and $(b^{n^2} y)$ to "conspire" and complicate the equidistribution properties of the sequence $(a^n x, b^{n^2} y)$.

If the transformation T_1 is ergodic but not totally ergodic, then further technical issues arise, but they are not hard to overcome. If T_1 is not ergodic, then it is possible to use its ergodic decomposition, and the previously established ergodic result per ergodic component, to deduce the result for T_1. Finally, if $f_2 \perp \mathcal{K}_{\mathrm{rat}}(T_2)$, we first use the previously established result to reduce matters to the case where the function f_1 is $\mathcal{K}_{\mathrm{rat}}(T_1)$-measurable, and then it becomes an easy matter to show that the averages of (34) converge to 0 in $L^2(\mu)$.

7.3. Proof of Theorem 1.3 modulo a convergence result.

We are going to derive Theorem 1.3 from the following result (that will be proved in the next subsection):

Proposition 7.3. *Let $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ be a system. Let $d_1, \ldots, d_\ell \in \mathbb{N}$ be distinct and $f_1, \ldots, f_\ell \in L^\infty(\mu)$. Suppose that $f_i \perp \mathcal{K}_{\mathrm{rat}}(T_i)$ for some $i = 1, \ldots, \ell$.*

Then for every $\varepsilon > 0$, there exists $r_0 \in \mathbb{N}$, such that for every $r \in r_0\mathbb{N}$, we have

$$(35) \qquad \lim_{N-M \to \infty} \left\| \frac{1}{N-M} \sum_{n=M}^{N-1} f_1(T_1^{(rn)^{d_1}} x) \cdot \ldots \cdot f_\ell(T_\ell^{(rn)^{d_\ell}} x) \right\|_{L^2(\mu)} \leq \varepsilon.$$

(The existence of the limit is given by Theorem 1.1.)

Remark. The conclusion should hold with $r_0 = 1$ and $\varepsilon = 0$, but we currently do not see how to show this.

We are also going to need the next inequality, it is proved by an appropriate application of Hölder's inequality:

Lemma 7.4 ([10])**.** *Let $\ell \in \mathbb{N}$, $(X, \mathcal{X}, \mu)$ be a probability space, $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_\ell$ be sub-σ-algebras of $\mathcal{X}$, and $f \in L^\infty(\mu)$ be non-negative.*

Then

$$\int f \cdot \mathbb{E}(f|\mathcal{X}_1) \cdot \mathbb{E}(f|\mathcal{X}_2) \cdot \ldots \cdot \mathbb{E}(f|\mathcal{X}_\ell) \, d\mu \geq \left(\int f \, d\mu \right)^{\ell+1}.$$

Proof of Theorem 1.3 assuming Proposition 7.3. Let $\varepsilon > 0$. It suffices to show that there exists $r \in \mathbb{N}$ such that

$$(36) \qquad \lim_{N-M \to \infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(A \cap T_1^{(rn)^{d_1}} A \cap \cdots \cap T_\ell^{(rn)^{d_\ell}} A) \geq \mu(A)^{\ell+1} - 2\varepsilon.$$

First we use Proposition 7.3 to choose $r_0 \in \mathbb{N}$ so that for every $r \in r_0\mathbb{N}$ we have the estimate (35) with $\varepsilon/2^\ell$ in place of ε. Next we choose a multiple r of r_0 such that for $i = 1, \ldots, \ell$ we have

$$(37) \qquad \|\mathbb{E}(\mathbf{1}_A|\mathcal{K}_r(T_i)) - \mathbb{E}(\mathbf{1}_A|\mathcal{K}_{\mathrm{rat}}(T_i))\|_{L^2(\mu)} \leq \frac{\varepsilon}{\ell}.$$

We claim that for this choice of r equation (36) holds. Indeed by (35) (with $\varepsilon/2^\ell$ in place of ε) we have that the limit in (36) is ε-close to the limit of the averages of

$$(38) \qquad \int \mathbf{1}_A \cdot T_1^{(rn)^{d_1}} \mathbb{E}(\mathbf{1}_A | \mathcal{K}_{\mathrm{rat}}(T_1)) \cdot \ldots \cdot T_\ell^{(rn)^{d_\ell}} \mathbb{E}(\mathbf{1}_A | \mathcal{K}_{\mathrm{rat}}(T_\ell)) \, d\mu.$$

Using (37) we easily conclude that the limit in (38) is ε close to the limit of the averages of

$$\int \mathbf{1}_A \cdot T_1^{(rn)^{d_1}} \mathbb{E}(\mathbf{1}_A|\mathcal{K}_r(T_1)) \cdot \ldots \cdot T_\ell^{(rn)^{d_\ell}} \mathbb{E}(\mathbf{1}_A|\mathcal{K}_r(T_\ell)) \, d\mu.$$

Since $T^r f = f$ for $\mathcal{K}_r(T)$-measurable functions f, the last expression is equal to

$$\int \mathbf{1}_A \cdot \mathbb{E}(\mathbf{1}_A | \mathcal{K}_r(T_1)) \cdot \ldots \cdot \mathbb{E}(\mathbf{1}_A | \mathcal{K}_r(T_\ell)) \; d\mu.$$

By Lemma 7.4, the last integral is greater or equal than $\mu(A)^{\ell+1}$. It follows that (36) holds and the proof is complete. $\qquad\square$

7.4. **Some equidistribution results.** In the next subsection we prove Proposition 7.3. A crucial step in the proof is an equidistribution result on nilmanifolds that we prove in this subsection. We start with a lemma.

Lemma 7.5. *Let $d, m_1, m_2 \in \mathbb{N}$ and $T \colon \mathbb{T}^{m_1} \to \mathbb{T}^{m_1}$ be an ergodic unipotent affine transformation. For $i = 1, \ldots, m_2$, let $u_i \in \mathbb{R}[t]$ be a polynomial divisible by t^{d+1}. Suppose that the sequence $(u(n))_{n \in \mathbb{N}}$, with values in $\mathbb{T}^{m_2}$, defined by*

$$u(n) = \big(u_1(n) \pmod 1, \ldots, u_{m_2}(n) \pmod 1\big)$$

is equidistributed on $\mathbb{T}^{m_2}$.

Then for $m_{\mathbb{T}^{m_1}}$-almost every $x \in \mathbb{T}^{m_1}$ the sequence $(T^{n^d} x, u(n))_{n \in \mathbb{N}}$ is equidistributed on $\mathbb{T}^{m_1} \times \mathbb{T}^{m_2}$. Furthermore, the set of full $m_{\mathbb{T}^{m_1}}$-measure can be chosen to depend only on the transformation T (so independently of the sequence $(u(n))_{n \in \mathbb{N}}$).

Proof. Suppose that $T \colon \mathbb{T}^{m_1} \to \mathbb{T}^{m_1}$ is defined by $Tx = Sx + b$ for some unipotent homomorphism S of $\mathbb{T}^{m_1}$ and $b \in \mathbb{T}^{m_1}$. We claim that the desired equidistribution property holds provided that x satisfies the following condition:

(39) $\qquad\qquad$ If $k_1 \cdot b + k_2 \cdot x = 0 \bmod 1$ for some $k_1, k_2 \in \mathbb{Z}^{m_1}$, then $k_2 = \mathbf{0}$.

This defines a set of full measure in $\mathbb{T}^{m_1}$ that depends only on the element $b \in \mathbb{T}^{m_1}$.

Let x_0 be any point in $\mathbb{T}^{m_1}$ that satisfies (39). Let χ be a non-trivial character of $\mathbb{T}^{m_1} \times \mathbb{T}^{m_2}$. Then $\chi = (\chi_1, \chi_2)$ for some characters χ_1 of $\mathbb{T}^{m_1}$ and χ_2 of $\mathbb{T}^{m_2}$, and either χ_1 or χ_2 is non-trivial. By Weyl's equidistribution theorem, in order to verify that the sequence $(T^{n^d} x_0, u(n))_{n \in \mathbb{N}}$ is equidistributed on $\mathbb{T}^{m_1} \times \mathbb{T}^{m_2}$, it suffices to show that

(40) $$\lim_{N-M \to \infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \chi_1(T^{n^d} x_0) \cdot \chi_2(u(n)) = 0.$$

If $\chi_1 = 1$, then (40) holds because, by assumption, the sequence $(u(n))_{n \in \mathbb{N}}$ is equidistributed on $\mathbb{T}^{m_2}$. Suppose now that $\chi_1 \neq 1$. Since $S \colon \mathbb{T}^{m_1} \to \mathbb{T}^{m_1}$ is unipotent, we have $(S-I)^{m_1} = 0$. For $n \geq m_1$, a straightforward induction shows that

$$T^n x_0 = \sum_{k=0}^{m_1-1} \binom{n}{k} (S-I)^k x_0 + \sum_{k=0}^{m_1-1} \binom{n}{k+1} (S-I)^k b .$$

Therefore, the sequence $(T^n x)_{n \in \mathbb{N}}$ is polynomial in n and

(41) $$T^n x_0 = x_0 + n\big((S-I)x_0 + b\big) + n^2 p(n)$$

for some polynomial $p \in \mathbb{R}[t]$.

Claim. $\chi_1((S-I)x_0 + b) = e(\alpha)$ *for some irrational number α.*

Suppose on the contrary that $\chi_1\big((S-I)x_0+b\big)$ is rational. After replacing χ_1 by some power of χ_1 we can assume that $\chi_1\big((S-I)x_0+b\big)=1$. We write $\chi_1(x)=e(k_1\cdot x)$ where k_1 is some non-zero element of $\mathbb{Z}^{m_1}$. Then $k_1\cdot\big((S-I)x_0+b\big)=0$ (mod 1), or equivalently,

$$(42) \qquad\qquad (k_1\cdot(S-I))\cdot x_0+k_1\cdot b=0 \pmod 1.$$

Combining (39) and (42) we get that $k_1\cdot(S-I)=0$ (mod 1). Using (42) again we get that $k_1\cdot b=0$ (mod 1). As a result, (42) holds for every $x\in\mathbb{T}^{m_1}$ in place of x_0, or equivalently, $\chi_1((S-I)x+b)=1$ for every $x\in\mathbb{T}^{m_1}$. Therefore, $\chi_1(Tx)=\chi_1(x)$. Since $\chi_1\neq 1$, this contradicts our assumption that the transformation T is ergodic. This completes the proof of the claim.

From (41) we conclude that

$$\chi_1(T^n x_0)=e(x_0+n\alpha+n^2 p(n))$$

for some irrational α.

Using this, and our assumption that all the polynomials $u_i(t)$ are divisible by t^{d+1}, we get that

$$\chi_1(T^{n^d}x_0)\cdot\chi_2(u(n))=e(x_0+n^d\alpha+n^{d+1}\tilde{p}(n))$$

for some polynomial $\tilde{p}\in\mathbb{R}[t]$. Since α is irrational, it follows from this identity and Weyl's equidistribution criterion that (40) holds. This completes the proof. $\qquad\square$

Lemma 7.6. *Let $X=G/\Gamma$ be a connected nilmanifold, $a\in G$ be an ergodic element, and $d\in\mathbb{N}$. Let $Y=H/\Delta$ be a nilmanifold, $(g(n)y)_{n\in\mathbb{N}}$ defined by $g(n)=a_1^{p_1(n)}\cdot\ldots\cdot a_\ell^{p_\ell(n)}$ be a polynomial sequence on Y, and suppose that the polynomials $p_1,\ldots,p_\ell$ are all divisible by t^{d+1}.*

Then there exists $r_0\in\mathbb{N}$ such that for m_X-almost every $x\in X$ we have: For every $r\in r_0\mathbb{N}$, the sequence $\big((a^{(rn)^d}x,g(rn)y)\big)_{n\in\mathbb{N}}$ is equidistributed on the set $X\times\mathrm{cl}_Y\{g(rn)y,n\in\mathbb{N}\}$. Furthermore, the set of full m_X-measure can be chosen to depend only on the element $a\in G$ (so independently of Y, y, and $g(n)$).

Remark. It is crucial for our subsequent applications that the full m_X-measure set of the lemma does not depend on the polynomial sequence $(g(n)y)_{n\in\mathbb{N}}$. It is for this reason that we require the polynomials $p_1,\ldots,p_\ell$ to be divisible by t^{d+1}.

Proof. The connected case. Suppose first that the set $\mathrm{cl}_Y\{g(n)y,n\in\mathbb{N}\}$ is connected. In this case we are going to show that $r_0=1$ works.

First, by part (ii) of Theorem 7.1, the set $\mathrm{cl}_Y\{g(n)y,n\in\mathbb{N}\}$ is a sub-nilmanifold of Y. Substituting this set for Y we can assume that $Y=\mathrm{cl}_Y\{g(n)y,n\in\mathbb{N}\}$. By part (ii) of Theorem 7.1, we have

$$(43)\qquad\qquad \textit{the sequence } (g(n)y) \textit{ is equidistributed in } Y.$$

(a) First we claim that it suffices to show

    (i) For m_X-almost every $x\in X$, where the set of full measure depends only on a, the sequence $\big((a^{n^d}x,g(n)y)\big)$ is equidistributed on the set $X\times Y$.

Indeed, since the nilmanifolds X and $X\times Y$ are connected, it follows by part (iii) of Theorem 7.1 that for every $r\in\mathbb{N}$ we have $\mathrm{cl}_Y\{g(rn)y,n\in\mathbb{N}\}=Y$, and for every $r\in\mathbb{N}$ and every x in the set defined in (ii), the sequence $\big((a^{(rn)^d}x,g(rn)y)\big)$ is equidistributed on the set $X\times Y$. This proves the claim.

(b) Next, we use the convergence criterion given in Theorem 7.2.

Let $A_X = G/([G_0, G_0]\Gamma)$ be the affine torus of X, $A_Y = H/([H_0, H_0]\Delta)$ be the affine torus of Y, and $\pi_{A_X} \colon X \to A_X$, $\pi_{A_Y} \colon Y \to A_Y$ be the corresponding natural projections. We first remark that $A_X \times A_Y$ is the affine torus of $X \times Y$, with projection $\pi_X \times \pi_Y$.

Since the sequence $(g(n)y)$ is equidistributed in Y, the projection of this sequence onto A_Y is equidistributed on A_Y. By Theorem 7.2, in order to show the required equidistribution property (i), it suffices to verify the following statement:

> (ii) For m_X-almost every $x \in X$, where the set of full measure depends only on a, the sequence
> $$\left(\left(a^{n^d} \pi_{A_X}(x), a_1^{p_1(n)} \cdot \ldots \cdot a_\ell^{p_\ell(n)} \pi_{A_Y}(y) \right) \right)$$
> is equidistributed on $A_X \times A_Y$.

This statement is the same as (i), with A_X substituted for X, A_Y substituted for Y. We remark that all the hypotheses of the lemma remain valid when we make this substitution.

Therefore, using the identification explained in Section 7.1, we can restrict, without loss of generality, to the case where $X = \mathbb{T}^{m_1}$ for some $m_1 \in \mathbb{N}$, the translation $T_a \colon x \mapsto ax$ on X is an ergodic unipotent affine transformation of $\mathbb{T}^{m_1}$, and where $Y = \mathbb{T}^{m_2}$ for some integer $m_2 \in \mathbb{N}$ and for $i = 1, \ldots, \ell$ the translation $T_{a_i} \colon y \mapsto a_i y$ on Y is a unipotent affine transformation of $\mathbb{T}^{m_2}$. Moreover, by (43), the sequence $(T_{a_1}^{p_1(n)} \cdot \ldots \cdot T_{a_\ell}^{p_\ell(n)} y)$ is equidistributed on $\mathbb{T}^{m_2}$.

Since the uniform distribution is not affected by translation, the statement (ii) can be rewritten in the following equivalent form:

> (iii) For $m_{\mathbb{T}^{m_1}}$-almost every $x \in \mathbb{T}^{m_1}$, where the set of full measure depends only on the transformation T_{a_1}, the sequence
> $$\left(\left(T_a^{n^d} x, T_{a_1}^{p_1(n)} \cdot \ldots \cdot T_{a_\ell}^{p_\ell(n)} y - y \right) \right)$$
> is equidistributed on $\mathbb{T}^{m_1} \times \mathbb{T}^{m_2}$.

(c) Define the sequence $(u(n))_{n \in \mathbb{N}}$ with values in $\mathbb{T}^{m_2}$ by
$$u(n) = T_{a_1}^{p_1(n)} \cdot \ldots \cdot T_{a_\ell}^{p_\ell(n)} y - y \ .$$

For $i = 1, \ldots, \ell$, since T_{a_i} is a unipotent affine transformation, $T_{a_i}^n y$ is given for every n by a formula similar to (41). Therefore, for $j = 1, \ldots, m_2$, each coordinate $u_j(n)$ of $u(n)$ is a polynomial in n with real coefficients and without a constant term. Moreover, since by hypothesis the polynomials $p_i(t)$ are divisible by t^{d+1}, all the polynomials $u_j(t)$ are divisible by t^{d+1}.

Hence, Lemma 7.5 is applicable and the statement (iii) is proved. This completes the proof of the result in the case where the set $\mathrm{cl}_Y\{g(n)y, n \in \mathbb{N}\}$ is connected.

The general case. Lastly we deal with the case where the set $\mathrm{cl}_Y\{g(n)y, n \in \mathbb{N}\}$ is not necessarily connected. By Theorem 7.1, there exists an $r_0 \in \mathbb{N}$ such that the set $\mathrm{cl}_Y\{g(r_0 n)y, n \in \mathbb{N}\}$ is connected. Substituting the sequence $(g(r_0 n)y)$ for $(g(n)y)$ and $a^{r_0^d}$ for a (which is again an ergodic element), the previous argument shows the advertised result for this value of $r_0 \in \mathbb{N}$. This completes the proof of the result in the general case. $\qquad \square$

We deduce from the previous lemma a result that is more suitable for our purposes:

Corollary 7.7. *Let $X = G/\Gamma$ be a nilmanifold, $a \in G$ be an ergodic element, $f \in \mathcal{C}(X)$ with $\mathbb{E}_{m_X}(f|\mathcal{K}_{\mathrm{rat}}(T_a)) = 0$, and $d, d_1 \ldots, d_\ell \in \mathbb{N}$ with $d < d_i$ for $i = 1, \ldots, \ell$. Suppose that $(u_{1,n})_{n \in \mathbb{N}}, \ldots, (u_{\ell,n})_{n \in \mathbb{N}}$, are finite step nilsequences.*

Then there exists $r_0 \in \mathbb{N}$ such that, for m_X-almost every $x \in X$, the following holds: For every $r \in r_0 \mathbb{N}$ we have

$$(44) \qquad \lim_{N-M \to \infty} \frac{1}{N-M} \sum_{n=M}^{N-1} f(a^{(rn)^d} x) \cdot u_{1,(rn)^{d_1}} \cdot \ldots \cdot u_{\ell,(rn)^{d_\ell}} = 0.$$

Furthermore, the set of full m_X-measure can be chosen to depend only on the element $a \in G$.

Proof. The connected case. Suppose first that the nilmanifold X is connected. Using an approximation argument we can assume that for $i = 1, \ldots, \ell$ the sequence $u_{i,n}$ is a basic finite step nilsequence. In this case, for $i = 1, \ldots, \ell$ there exist nilmanifolds $X_i = G_i/\Gamma_i$, elements $a_i \in G_i$, and functions $f_i \in \mathcal{C}(X_i)$ such that $u_{i,n} = f_i(a_i^n \Gamma_i)$. We define $\tilde{G} = G_1 \times \cdots \times G_\ell$, $\tilde{\Gamma} = \Gamma_1 \times \cdots \times \Gamma_\ell$, and $\tilde{X} = X_1 \times \cdots \times X_\ell = \tilde{G}/\tilde{\Gamma}$. Let $(g(n))_{n \in \mathbb{N}}$ be the polynomial sequence in $\tilde{G}$ given by $g(n) = (a_1^{p_1(n)}, \ldots, a_\ell^{p_\ell(n)})$ for every n.

Lemma 7.6 gives that there exists $r_0 \in \mathbb{N}$ such that for m_X-almost every $x \in X$ we have: For every $r \in r_0 \mathbb{N}$, the sequence $(a^{(rn)^d} x, g(rn)\tilde{\Gamma})$ is equidistributed on the nilmanifold $X \times Y$ where $Y = \mathrm{cl}_{\tilde{X}}\{g(rn)\tilde{\Gamma}, n \in \mathbb{N}\}$.

Therefore, for every $f \in \mathcal{C}(X)$ and $F \in \mathcal{C}(\tilde{X})$ we have

$$\lim_{N-M \to \infty} \frac{1}{N-M} \sum_{n=M}^{N-1} f(a^{(rn)^d} x) \cdot F(g(rn)\tilde{\Gamma}) = \int f(x) \, dm_X(x) \cdot \int F(\tilde{x}) \, dm_{\tilde{X}}(\tilde{x}).$$

Letting $F = f_1 \cdot \ldots \cdot f_\ell$, and using that $\int f \, dm_X = 0$, we get the advertised identity. This completes the proof in the case where the nilmanifold X is connected.

The general case. Let X_0 be the connected component of the nilmanifold X. Since a is an ergodic element, there exists $k \in \mathbb{N}$ such that the nilmanifold X is the disjoint union of the sub-nilmanifolds $X_i = a^i X_0$, $i = 1, \ldots, k$, and a^k acts ergodically on each X_i. Furthermore, since $\mathbb{E}_{m_X}(f|\mathcal{K}_{\mathrm{rat}}(T_a)) = 0$, we have $\int f \, dm_{X_i} = 0$ for $i = 1, \ldots, k$. For $i = 1, \ldots, k$, we can apply the previously established "connected result", for the translation a^{k^d} in place of a, acting (ergodically) on the connected sub-nilmanifolds X_i, and the nilsequences $(u_{j,k^{d_j} n})$ in place of $(u_{j,n})$, $j = 1, \ldots, \ell$. We get that there exist $r_i \in \mathbb{N}$ such that for every $r \in kr_i \mathbb{N}$ equation (44) holds for m_{X_i}-almost every $x \in X_i$. It follows that if $r_0 = k \prod_{i=1}^{k} r_i$, then for every $r \in r_0 \mathbb{N}$ equation (44) holds for m_X-almost every $x \in X$. This completes the proof in the general case. $\qquad \square$

7.5. Proof of the convergence result (Proposition 7.3).
In this section we prove Proposition 7.3 by induction on the number of transformations involved. The key ingredient in the proof of the inductive step is the following special case of Proposition 7.3:

Lemma 7.8. *Let $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ be a system. Let $d_1, \ldots, d_\ell \in \mathbb{N}$ be distinct and suppose that $d_1 < d_i$ for $i = 2, \ldots, \ell$. Let $f_1, \ldots, f_\ell \in L^\infty(\mu)$ and suppose that $f_1 \perp \mathcal{K}_{\mathrm{rat}}(T_1)$.*

Then for every $\varepsilon > 0$, there exists $r_0 \in \mathbb{N}$, such that for every $r \in r_0 \mathbb{N}$, we have

$$\lim_{N-M \to \infty} \left\| \frac{1}{N-M} \sum_{n=M}^{N-1} f_1(T_1^{(rn)^{d_1}} x) \cdot \ldots \cdot f_\ell(T_\ell^{(rn)^{d_\ell}} x) \right\|_{L^2(\mu)} \leq \varepsilon.$$

Proof. Let $\varepsilon > 0$. Without loss of generality we can assume that all the functions involved are bounded by 1. From Theorem 1.2 we have that there exists $k \in \mathbb{N}$, depending only on $\max(d_1, \ldots, d_\ell)$ and ℓ, such that if $f_i \perp \mathcal{Z}_{k,T_i}$ for some $i = 1, \ldots, \ell$, then the corresponding multiple ergodic averages converge to 0 in $L^2(\mu)$. Therefore, we can assume that $f_i \in L^\infty(\mathcal{Z}_{k,T_i}, \mu)$ for $i = 1, \ldots, \ell$. Then Proposition 3.1 shows that for $i = 2, \ldots, \ell$ there exist functions $\tilde{f}_i$, with L^∞-norm bounded by 1, that satisfy

(i) $\tilde{f}_i \in L^\infty(\mathcal{Z}_{k,T_i}, \mu)$ and $\left\| f_i - \tilde{f}_i \right\|_{L^2(\mu)} \leq \varepsilon/(2\ell+2)$;

(ii) for every $r \in \mathbb{N}$ and $x \in X$ the sequence $(\tilde{f}_i(T_i^{n^{d_i}} x))_{n \in \mathbb{N}}$ is a $(d_i k)$-step nilsequence.

An easy computation then shows that in order to prove the announced claim, it suffices to show the following: If $f_1 \in L^\infty(\mathcal{Z}_{k,T_1}, \mu)$ and $f_1 \perp \mathcal{K}_{\mathrm{rat}}(T_1)$, then there exists $r_0 \in \mathbb{N}$ such that for every $r \in r_0 \mathbb{N}$ we have

(45) $$\lim_{N-M \to \infty} \left\| \frac{1}{N-M} \sum_{n=M}^{N-1} f_1(T_1^{(rn)^{d_1}} x) \cdot \tilde{f}_2(T_2^{(rn)^{d_2}} x) \cdot \ldots \cdot \tilde{f}_\ell(T_\ell^{(rn)^{d_\ell}} x) \right\|_{L^2(\mu)} \leq \frac{\varepsilon}{2}.$$

The ergodic case. Suppose first that the transformation T_1 is ergodic. Since $f_1 \in L^\infty(\mathcal{Z}_{k,T_1}, \mu)$, after using an appropriate conjugation we can assume that T_1 is an inverse limit of nilsystems. Furthermore, after using an approximation argument we can assume that $T_1 = T_a$ where a is an ergodic rotation on a nilmanifold $X = G/\Gamma$, and $f_1 \in C(X)$, while still maintaining our assumption that $f_1 \perp \mathcal{K}_{\mathrm{rat}}(T_1)$. (If $f \perp \mathcal{D}$ where $\mathcal{D}$ is any sub-σ-algebra of $\mathcal{X}$, and g is such that $\|f - g\|_{L^1(\mu)} \leq \varepsilon/2$, then $\|\mathbb{E}(g|\mathcal{D})\|_{L^1(\mu)} \leq \varepsilon/2$. Therefore, $\|f - \tilde{g}\|_{L^1(\mu)} \leq \varepsilon$ where $\tilde{g} = g - \mathbb{E}(g|\mathcal{D})$, and $\mathbb{E}(\tilde{g}|\mathcal{D}) = 0$.) In this case, combining property (ii) above and Corollary 7.7, we get that there exists $r_0 \in \mathbb{N}$ such that for every $r \in r_0 \mathbb{N}$ the averages (45) converge to 0 for m_X-almost every $x \in X$, and as a result in $L^2(m_X)$. This completes the proof of (45) in the case where the transformation T_1 is ergodic.

The general case. Suppose now that the transformation T_1 is not necessarily ergodic. Let $\mu = \int \mu_x \, d\mu$ be the ergodic decomposition of μ with respect to the transformation T_1. Since $f_1 \in L^\infty(\mathcal{Z}_{k,T_1,\mu}, \mu)$, Corollary 3.3 shows that for μ-almost every $x \in X$ we have $f_1 \in L^\infty(\mathcal{Z}_{k,T_1,\mu_x}, \mu_x)$. Furthermore, since $\mathbb{E}_\mu(f_1 | \mathcal{K}_{\mathrm{rat}}(T, \mu)) = 0$, we have for μ-almost every $x \in X$ that $\mathbb{E}_{\mu_x}(f_1 | \mathcal{K}_{\mathrm{rat}}(T, \mu_x)) = 0$.

For every $r_0 \in \mathbb{N}$ we define the μ-measurable set

$$X_{r_0} = \big\{ x \in X \colon (45) \text{ holds for every } r \in r_0 \mathbb{N}, \text{ with } \mu_x \text{ in place of } \mu, \text{ and } \varepsilon/2 \text{ in place of } \varepsilon \big\}.$$

Notice that when we previously established the "ergodic case", we did not use the invariance of the measure μ under the transformations T_i for $i \neq 1$; we merely used the fact that for $i = 2, \ldots, \ell$, for μ-almost every $x \in X$, the sequences $(\tilde{f}_i(T_i^n x))_{n \in \mathbb{N}}$ are k-step nilsequences.

Hence, we can use the previously established "ergodic result" for μ-almost every measure μ_x, and conclude that

$$\mu\left(\bigcup_{r_0 \in \mathbb{N}} X_{r_0}\right) = 1.$$

Also, we clearly have $X_r \subset X_s$ if r divides s. It follows that there exists $r_0 \in \mathbb{N}$ such that

$$\mu(X_{r_0}) \geq 1 - \varepsilon/4.$$

As a direct consequence, for this choice of r_0, equation (45) holds for every $r \in r_0\mathbb{N}$. This completes the proof. $\qquad\square$

We are now ready to prove Proposition 7.3.

Proof of Proposition 7.3. Without loss of generality we can assume that $d_1 < d_i < d_\ell$ for $i = 2, \ldots, \ell - 1$ and $\|f_i\|_{L^\infty(\mu)} \leq 1$ for $i = 1, \ldots, \ell$.

We are going to use induction on the number of transformations ℓ. For $\ell = 1$ the statement is known (Chapter 3 in [18]) and in fact it holds with $r_0 = 1$ and $\varepsilon = 0$. Suppose that $\ell \geq 2$, and the statement holds for $\ell - 1$ transformations. We are going to show that it holds for ℓ transformations. Namely, we are going to show that if $f_i \perp \mathcal{K}_{\mathrm{rat}}(T_i)$ for some $i \in \{1, \ldots, \ell\}$, then for every $\varepsilon > 0$, there exists $r_0 \in \mathbb{N}$, such that for every $r \in r_0\mathbb{N}$ we have

$$\lim_{N-M\to\infty} \left\| \frac{1}{N-M} \sum_{n=M}^{N-1} f_1(T_1^{(rn)^{d_1}} x) \cdot \ldots \cdot f_\ell(T_\ell^{(rn)^{d_\ell}} x) \right\|_{L^2(\mu)} \leq \varepsilon.$$

Let $\varepsilon > 0$. If $f_1 \perp \mathcal{K}_{\mathrm{rat}}(T_1)$, then the result follows from Lemma 7.8. So we can assume that $f_i \perp \mathcal{K}_{\mathrm{rat}}(T_i)$ for some $i \in \{2, \ldots, \ell\}$. By Lemma 7.8 we can assume that the function f_1 is $\mathcal{K}_{\mathrm{rat}}(T_1)$-measurable. Furthermore, using a standard approximation argument we can assume that the function f_1 is $\mathcal{K}_{r_1}(T_1)$-measurable for some $r_1 \in \mathbb{N}$. Since for every $r \in r_1\mathbb{N}$ we have $T^r f_1 = f_1$, it remains to find $r_2 \in r_1\mathbb{N}$, such that for every $r \in r_2\mathbb{N}$ we have

$$\lim_{N-M\to\infty} \left\| \frac{1}{N-M} \sum_{n=M}^{N-1} f_2(T_2^{(rn)^{d_1}} x) \cdot \ldots \cdot f_\ell(T_\ell^{(rn)^{d_\ell}} x) \right\|_{L^2(\mu)} \leq \varepsilon.$$

Such an integer r_2 exists from the induction hypothesis. This completes the induction and the proof. $\qquad\square$

Appendix A. Some "simple" proofs of special cases of the main results

It turns out that Theorem 1.2 can be strengthened, and the proof of Theorems 1.1, 1.2, and 1.3, can be greatly simplified in some interesting special cases, namely when $\ell = 2$ and one of the two polynomials is linear. Such a simplification is feasible because of the nature of the averages involved; it turns out to be possible to get simple characteristic factors by using a variation of van der Corput's Lemma, and then appealing to a known result from [15]. We take the opportunity in this section to give these simple arguments. Hopefully, the non-persistent reader, that does not want to embark to the details of the more complicated proofs of our main results, will benefit from the proofs of the special cases given here.

The key ingredient in the proofs is the following result:

Theorem A.1 ([15])**.** *Let $(X, \mathcal{X}, \mu, T)$ be a system and suppose that the integer polynomials $1, p, q$ are linearly independent. Let $f, g \in L^\infty(\mu)$ and suppose that $f \perp \mathcal{K}_{\mathrm{rat}}(T)$ or $g \perp \mathcal{K}_{\mathrm{rat}}(T)$. Then*

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} f(T^{p(n)}x) \cdot g(T^{q(n)}x) = 0$$

where the convergence takes place in $L^2(\mu)$.

Remark. The proof in [15] is given for ergodic systems, but the announced result follows directly from this, since $f \perp \mathcal{K}_{\mathrm{rat}}(T, \mu)$ implies that $f \perp \mathcal{K}_{\mathrm{rat}}(T, \mu_x) = 0$ for μ-almost every $x \in X$, where as usual, $\mu = \int \mu_x \, d\mu(x)$ is the ergodic decomposition of μ.

We are also going to use the following variation of the classical elementary lemma of van der Corput. Its proof is a straightforward modification of the one given in [5].

Lemma A.2. *Let $\{v_{N,n}\}_{N,n\in\mathbb{N}}$ be a bounded sequence of vectors in a Hilbert space. For every $h \in \mathbb{N}$ we set*

$$b_h = \overline{\lim}_{N\to\infty} \left| \frac{1}{N} \sum_{n=1}^{N} <v_{N,n+h}, v_{N,n}> \right|.$$

Suppose that

$$\lim_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} b_h = 0.$$

Then

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} v_{N,n} \right\| = 0.$$

We start with the following strengthening of Theorem 1.2 in our particular setup:

Theorem A.3. *Let $(X, \mathcal{X}, \mu, T, S)$ be a system. Let $f, g \in L^\infty(\mu)$ and suppose that either $f \perp \mathcal{K}_{\mathrm{rat}}(T)$ or $g \perp \mathcal{K}_{\mathrm{rat}}(S)$.*
Then for every polynomial $p \in \mathbb{Z}[t]$ with $\deg(p) \geq 2$ we have

$$(46) \qquad \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} f(T^n x) \cdot g(S^{p(n)}x) = 0$$

where the convergence takes place in $L^2(\mu)$.

Proof. Suppose first that $\mathbb{E}(g| \mathcal{K}_{\mathrm{rat}}(S)) = 0$. It suffices to show that for every sequence of intervals $(I_N)_{N\in\mathbb{N}}$ with length increasing to infinity, the averages in n over the intervals I_N of

$$\int h_N(x) \cdot f(T^n x) \cdot g(S^{p(n)}x) \, d\mu$$

converge to 0, where $h_N(x) = \frac{1}{|I_N|} \sum_{n\in I_N} f(T^n x) \cdot g(S^{p(n)}x)$. Equivalently, it suffices to show that the averages over the intervals I_N of

$$\int f(x) \cdot h_N(T^{-n}x) \cdot g(S^{p(n)}T^{-n}x) \, d\mu$$

converge to 0. Using the Cauchy-Schwarz inequality it suffices to show that the averages over the intervals I_N of

$$h_N(T^{-n}x) \cdot g(S^{p(n)}T^{-n}x)$$

converge to 0 in $L^2(\mu)$. By Lemma A.2 it suffices to show that for every $m \in \mathbb{N}$ the averages in n over the intervals I_N of

$$\int h_N(T^{-n}x) \cdot g(S^{p(n)}T^{-n}x) \cdot h_N(T^{-(n+m)}x) \cdot g(S^{p(n+m)}T^{-(n+m)}x) \, d\mu$$

converge to 0. We compose with the transformation T^n and use the Cauchy-Schwarz inequality. It suffices to show that for every $m \in \mathbb{N}$ the averages in n over the intervals I_N of

$$g(S^{p(n)}x) \cdot g(T^{-m}S^{p(n+m)}x)$$

converge to 0 in $L^2(\mu)$. Since $\deg(p) \geq 2$, for every $m \in \mathbb{N}$ the polynomials $1, p(n), p(n+m)$ are linearly independent. Since $g \perp \mathcal{K}_{\mathrm{rat}}(S)$, Theorem A.1 verifies that the last identity holds.

It remains to show that if $f \perp \mathcal{K}_{\mathrm{rat}}(T)$, then the averages over the intervals I_N of

$$f(T^n x) \cdot g(S^{p(n)}x)$$

converge to 0 in $L^2(\mu)$. Using the previously established property we get that the above limit remains unchanged if we replace the function g with the function $\mathbb{E}(g|\mathcal{K}_{\mathrm{rat}}(S))$. Furthermore, using an approximation argument and linearity, we can assume that $Sg = e(r)g$ for some $r \in \mathbb{Q}$. In this case, it suffices to show that the averages over the intervals I_N of

$$f(T^n x) \cdot e(rp(n))$$

converge to 0 in $L^2(\mu)$. Using the spectral theorem for unitary operators it suffices to show that for every $r \in \mathbb{Q}$ we have

$$(47) \qquad \lim_{N \to \infty} \left\| \frac{1}{|I_N|} \sum_{n \in I_N} e(nt + rn^2) \right\|_{L^2(\sigma_f(t))} = 0$$

where σ_f denotes the spectral measure of the function f. Since $f \perp \mathcal{K}_{\mathrm{rat}}(T)$, the measure σ_f has no rational point masses. Furthermore, as is well known, for t irrational the averages in (47) converge to 0 pointwise. Combining these two facts, and using the bounded convergence theorem, we deduce that (47) holds. This completes the proof. $\qquad \square$

We deduce the following special case of Theorem 1.1:

Theorem A.4. *Let* $(X, \mathcal{X}, \mu, T, S)$ *be a system and* $f, g \in L^\infty(\mu)$. *Let* $p \in \mathbb{Z}[t]$ *with* $\deg(p) \geq 2$. *Then the limit*

$$\lim_{N-M \to \infty} \frac{1}{N-M} \sum_{n=M}^{N-1} f(T^n x) \cdot g(S^{p(n)}x)$$

exists in $L^2(\mu)$.

Proof. By Theorem A.3 we can assume that the function f is $\mathcal{K}_{\mathrm{rat}}(T)$-measurable and the function g is $\mathcal{K}_{\mathrm{rat}}(S)$-measurable. Furthermore using an approximation argument we can assume that $T^r f = f$ and $T^r g = g$ for some $r \in \mathbb{N}$. In this case the result is obvious. $\qquad \square$

As a corollary we get an short proof for weak convergence of some multiple ergodic averages recently studied by T. Austin in [4] (where strong convergence was proven when $p(n) = n^2$).

Corollary A.5. *Let $(X, \mathcal{X}, \mu, T, S)$ be a system and $f, g \in L^\infty(\mu)$. Let $p \in \mathbb{Z}[t]$ with $\deg(p) \geq 2$.*
Then the averages

$$(48) \qquad \frac{1}{N-M} \sum_{n=M}^{N-1} f(T^{p(n)}x) \cdot g(T^{p(n)}S^n x)$$

converge weakly in $L^2(\mu)$ as $N - M \to \infty$. Furthermore, the limit is 0 if either $g \perp \mathcal{K}_{\mathrm{rat}}(S)$ or $f \perp (\mathcal{K}_{\mathrm{rat}}(T) \vee \mathcal{K}_{\mathrm{rat}}(S))$.

Proof. Notice that for every $h \in L^\infty(\mu)$ the averages of

$$\int h(x) \cdot f(T^{p(n)}x) \cdot g(T^{p(n)}S^n x) \; d\mu$$

are equal to the averages of

$$(49) \qquad \int f(x) \cdot h(T^{-p(n)}x) \cdot g(S^n x) \; d\mu.$$

Theorem A.4 shows that the averages of (49) converge, therefore the averages (48) converge weakly. Furthermore, Theorem A.3 shows that the averages of (49) converge to 0 if either $g \perp \mathcal{K}_{\mathrm{rat}}(S)$ or $h \perp \mathcal{K}_{\mathrm{rat}}(T)$, and as a consequence they converge to 0 if $f \perp (\mathcal{K}_{\mathrm{rat}}(T) \vee \mathcal{K}_{\mathrm{rat}}(S))$. Therefore, if $g \perp \mathcal{K}_{\mathrm{rat}}(S)$ or $f \perp (\mathcal{K}_{\mathrm{rat}}(T) \vee \mathcal{K}_{\mathrm{rat}}(S))$, then the averages of (49) converge weakly to 0. This completes the proof. $\square$

Finally we establish the following result:

Theorem A.6. *Let $(X, \mathcal{X}, \mu, T, S)$ be a system and $A \in \mathcal{X}$. Let $p \in \mathbb{Z}[t]$ with $\deg(p) \geq 2$ and $p(0) = 0$.*
Then for every positive integer $k \geq 2$ and $\varepsilon > 0$ the set

$$\{n \in \mathbb{N} \colon \mu(A \cap T^{-n}A \cap S^{-p(n)}A) > \mu(A)^3 - \varepsilon\}$$

has bounded gaps.

Proof. Let $\varepsilon > 0$. There exists $r \in \mathbb{N}$ such that
$$(50)$$
$$\|\mathbb{E}(\mathbf{1}_A|\mathcal{K}_r(T)) - \mathbb{E}(\mathbf{1}_A|\mathcal{K}_{\mathrm{rat}}(T))\|_{L^2(\mu)} \leq \varepsilon/3, \quad \|\mathbb{E}(\mathbf{1}_A|\mathcal{K}_r(S)) - \mathbb{E}(\mathbf{1}_A|\mathcal{K}_{\mathrm{rat}}(S))\|_{L^2(\mu)} \leq \varepsilon/3.$$

It suffices to show that

$$\lim_{N-M \to \infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(A \cap T^{-rn}A \cap S^{-p(rn)}A) \geq \mu(A)^3 - \varepsilon.$$

Using a straightforward modification of Theorem A.3, where T^n is replaced with T^{rn}, we see that the previous limit is equal to the limit of the averages of

$$\int \mathbf{1}_A \cdot T^{-rn}\mathbb{E}(\mathbf{1}_A|\mathcal{K}_{\mathrm{rat}}(T)) \cdot S^{-p(rn)}\mathbb{E}(\mathbf{1}_A|\mathcal{K}_{\mathrm{rat}}(S)) \; d\mu.$$

Using (50) we see that the last limit is ε-close to the limit of the averages of

$$\int \mathbf{1}_A \cdot T^{-rn}\mathbb{E}(\mathbf{1}_A|\mathcal{K}_r(T)) \cdot S^{-p(rn)}\mathbb{E}(\mathbf{1}_A|\mathcal{K}_r(S)) \; d\mu.$$

Since $T^r f = f$ for $\mathcal{K}_r(T)$-measurable functions f, $S^r f = f$ for $\mathcal{K}_r(S)$-measurable functions f, and $r|p(rn)$ for every $n \in \mathbb{N}$ (since $p(0) = 0$), the last limit is equal to

$$\int \mathbf{1}_A \cdot \mathbb{E}(\mathbf{1}_A|\mathcal{K}_r(T)) \cdot \mathbb{E}(\mathbf{1}_A|\mathcal{K}_r(S)) \, d\mu.$$

By Lemma 7.4, the last integral is greater or equal than $\mu(A)^3$, completing the proof. $\square$

References

[1] I. Assani. Pointwise convergence of ergodic averages along cubes. *J. Analyse Math.* **110** (2010), 241–269.

[2] T. Austin. On the norm convergence of non-conventional ergodic averages. *Ergodic Theory Dynam. Systems* **30** (2010), 321–338.

[3] T. Austin. Pleasant extensions retaining algebraic structure, I. *Preprint.* `arXiv:0905.0518v4`.

[4] T. Austin. Pleasant extensions retaining algebraic structure, II. *Preprint.* `arXiv:0910.0907v3`.

[5] V. Bergelson. Weakly mixing PET. *Ergodic Theory Dynam. Systems* **7** (1987), 337–349.

[6] V. Bergelson. Ergodic Ramsey Theory – an update, Ergodic Theory of $\mathbb{Z}^d$-actions (edited by M. Pollicott and K. Schmidt). *London Math. Soc. Lecture Note Series* **228** (1996), 1–61.

[7] V. Bergelson, B. Host, B. Kra, with an Appendix by I. Ruzsa. Multiple recurrence and nilsequences. *Inventiones Math.* **160** (2005), 261-303.

[8] V. Bergelson, A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *Journal Amer. Math. Soc.* **9** (1996), 725–753.

[9] V. Bergelson, A. Leibman. Failure of Roth theorem for solvable groups of exponential growth. *Ergodic Theory Dynam. Systems* **24** (2004), no. 1, 45–53.

[10] Q. Chu. Multiple recurrence for two commuting transformations. To appear *Ergodic Theory Dynam. Systems* Available at arXiv:0912.3381

[11] J.-P. Conze, E. Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France* **112** (1984), 143–175.

[12] J.-P. Conze, E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *Publications de l'Institut de Recherche de Mathématiques de Rennes, Probabilités*, 1987.

[13] J.-P. Conze, E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *C. R. Acad. Sci. Paris, Série I* **306** (1988), 491–493.

[14] N. Frantzikinakis. Multiple ergodic averages for three polynomials and applications. *Trans. Amer. Math. Soc.* **360** (2008), no. 10, 5435-5475.

[15] N. Frantzikinakis, B. Kra. Polynomial averages converge to the product of the integrals. *Isr. J. Math.* **148** (2005), 267-276.

[16] N. Frantzikinakis, B. Kra. Convergence of multiple ergodic averages for some commuting transformations. *Ergodic Theory Dynam. Systems* **25** (2005), 799–809.

[17] N. Frantzikinakis, B. Kra. Ergodic averages for independent polynomials and applications. *J. London Math. Soc.* **74** (2006), no. 1, 131–142.

[18] H. Furstenberg. Recurrence in ergodic theory and combinatorial number theory. *Princeton University Press*, Princeton, (1981).

[19] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.* **31** (1977), 204-256.

[20] H. Furstenberg, Y. Katznelson. An ergodic Szemerédi theorem for commuting transformations. *J. Analyse Math.* **34** (1978), 275–291.

[21] H. Furstenberg, B. Weiss. A mean ergodic theorem for $\frac{1}{N}\sum_{n=1}^{n} f(T^n x)g(T^{n^2} x)$. *Convergence in Ergodic Theory and Probability*, Eds.:Bergelson, March, Rosenblatt. Walter de Gruyter & Co, Berlin, New York (1996), 193–227.

[22] B. Host. Ergodic seminorms for commuting transformations and applications. *S*tudia Math. **195** (1) (2009), 31–49.

[23] B. Host, B. Kra. Non-conventional ergodic averages and nilmanifolds. *Ann. Math.* **161** (2005), 397-488.

[24] B. Host, B. Kra. Convergence of polynomial ergodic averages. *Israel J. of Math.* **149** (2005), 1–20.

[25] B. Host, B. Kra. Uniformity seminorms on l^∞ and spplications. *J. Analyse Math.* **108** (2009), 219–276.
[26] B. Host, B. Kra, A. Maass. Nilsequences and a topological structure theorem. *Advances in Math.* **224** (2010), 103-129.
[27] M. Johnson. Convergence of polynomial ergodic averages for some commuting transformations. To appear, *Illinois J. Math.*
[28] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of rotations of a nilmanifold. *Ergodic Theory Dynam. Systems* **25** (2005), no. 1, 201–213.
[29] E. Lesigne. Équations fonctionnelles, couplages de produits gauches et théorèmes ergodiques pour mesures diagonales. *Bull. Soc. Math. France* **121** (1993), no. 3, 315–351.
[30] D.J. Rudolph. Eigenfunctions of $T \times S$ and the Conze-Lesigne algebra. *Ergodic Theory and its Connections with Harmonic Analysis*, Eds.:Petersen/Salama, Cambridge University Press, New York (1995), 369–432.
[31] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory Dynam. Systems* **28** (2008), no. 2, 657–688.
[32] H. Towsner. Convergence of diagonal ergodic averages. *Ergodic Theory Dynam. Systems* **29** (2009), 1309–1326.
[33] Q. Zhang. On convergence of the averages $\frac{1}{N} \sum_{n=1}^{N} f_1(R^n x) f_2(S^n x) f_3(T^n x)$. *Monatsh. Math.* **122** (1996), 275-300.
[34] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.* **20** (2007), 53–97.

(Qing Chu) Université Paris-Est Marne-la-Vallée, Laboratoire d'analyse et de mathématiques appliquées, UMR CNRS 8050, 5 Bd Descartes, 77454 Marne la Vallée Cedex, France
E-mail address: `qing.chu@univ-mlv.fr`

(Nikos Frantzikinakis) University of Crete, Department of mathematics, Knossos Avenue, Heraklion 71409, Greece
E-mail address: `frantzikinakis@gmail.com`

(Bernard Host) Université Paris-Est Marne-la-Vallée, Laboratoire d'analyse et de mathématiques appliquées, UMR CNRS 8050, 5 Bd Descartes, 77454 Marne la Vallée Cedex, France
E-mail address: `bernard.host@univ-mlv.fr`